



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

ALGUNOS MÉTODOS DE ESTIMACIÓN Y SIMULACIÓN PARA PROCESOS DE DIFUSIÓN

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

MATEMÁTICO

PRESENTA:

CÉSAR AUGUSTO PÉREZ ROSAS

TUTOR

DR. FERNANDO BALTAZAR LARIOS

CIUDAD UNIVERSITARIA, CD. MX. 2020





Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

1. Datos del alumno.

Pérez

Rosas

César Augusto

Universidad Nacional Autónoma de México

Facultad de Ciencias

Matemáticas

308342581

2. Datos del asesor.

Baltazar

Larios

Fernando

3. Datos de la tesis

Algunos métodos de estimación y simulación para procesos de difusión

94p.

2020

Agradecimientos

Estoy profundamente agradecido con mis padres, mi mamá Rosy por su bondad, apoyo y su gran amor que siempre me ha entregado, mi padre, por su responsabilidad, apoyo, fortaleza emocional y motivación que me ha entregado siempre, mi hermana porque siempre ha sido mi compañía y a mi Ama, que ha sido mi equilibrio y mejor amiga.

Agradezco a mi tutor Fernando y a mis sinodales Ramsés, Arnaud, Carlos Erwin y Eduardo Selim por sus valiosas observaciones y aportaciones para la culminación de este trabajo.

Agradezco a mis mentores y amigos, que son demasiados y nunca terminaría, que me han visto crecer y han estado ahí siempre.

Especialmente agradezco a mi mentor y ahora amigo Jaime por darme la oportunidad de trabajar con él en una de las profesiones más lindas que existen.

Este trabajo fue realizado gracias al apoyo del Programa de Apoyo para la Innovación y Mejoramiento de la Enseñanza(PAPIME) de la UNAM PE102618.

Índice general

Agradecimientos	III
1 Preliminares de procesos estocásticos	3
1.1 Definiciones	3
2 Cálculo de Itô	7
2.1 Construcción de la integral de Itô	7
2.1.1 Integral de Itô	11
2.2 Fórmula de Itô	16
2.2.1 Transformación de Lamperti	19
3 Ecuaciones diferenciales estocásticas	21
3.1 Introducción	21
3.2 Un resultado de existencia y unicidad para ecuaciones diferenciales estocásticas	23
3.3 Soluciones fuertes y débiles	26
4 Procesos de difusión.	28
4.1 Definiciones	28
4.2 Procesos de difusión como procesos de Markov	29
4.3 Ecuaciones de Kolmogorov	30
4.3.1 Ecuación de Kolmogorov hacia adelante	30
4.3.2 Ecuación de Kolmogorov hacia atrás	30
4.4 Procesos de difusión como soluciones a ecuaciones diferenciales estocásticas	31
4.5 El generador infinitesimal de una difusión	32
5 Algunas familias paramétricas de procesos estocásticos	34
5.1 Proceso de Ornstein-Uhlenbeck	34
5.2 Modelo de Black-Scholes-Merton o movimiento Browniano geométrico	36

5.3	Modelo de Cox-Ingersoll-Ross	37
6	Aproximación de soluciones para ecuaciones diferenciales estocásticas vía simulación	39
6.1	Órdenes de Convergencia	39
6.2	Esquema de Euler	40
6.3	Esquema de Milstein	44
7	Inferencia para procesos de Difusión	49
7.1	Estimación máxima verosímil	50
7.1.1	Ornstein-Uhlenbeck	51
7.1.2	Modelo de Black-Scholes-Merton	55
7.1.3	Proceso de Cox-Ingersoll-Ross	57
7.2	Estimación máxima verosímil por el método de Euler-Maruyama	58
7.2.1	Ornstein-Uhlenbeck	60
7.2.2	Proceso de Cox-Ingersoll-Ross	61
7.3	Expansión de la verosimilitud via polinomios de Hermite	63
7.3.1	Aproximación de la función de transición por polinomios de Hermite	69
8	Conclusiones	75
A	Simulación en R	78
A.1	Estimación máxima verosímil	78
A.1.1	Estimación máxima verosímil para el proceso de Ornstein-Uhlenbeck	78
A.1.2	Estimación máxima verosímil para el modelo de Black-Scholes-Merton	79
A.1.3	Estimación máxima verosímil para n trayectorias del proceso de Ornstein-Uhlenbeck	81
A.2	Estimación máxima verosímil por el método de Euler-Maruyama	85
A.2.1	Estimación para el proceso de Ornstein-Uhlenbeck	85
A.2.2	Estimación para el proceso de Cox-Ingersoll-Ross	87

Prefacio

La presente tesis tiene como objetivo mostrar métodos de estimación y simulación para algunos de los procesos estocásticos más estudiados en el mundo financiero. Esta clase de procesos estocásticos son conocidos como procesos de difusión o difusiones que pertenecen a una clase mucho más grande de procesos, a saber, procesos de Markov, estos procesos tienen asociada una probabilidad de transición la cual es la pieza angular para los métodos de estimación presentados en este trabajo. Una forma de ver a los procesos de difusión es como la solución a una ecuación diferencial estocástica, por lo que en el Capítulo 2 se presentará de una forma no exhaustiva la construcción de la integral estocástica (integral de Itô), análoga a la construcción de la integral de Lebesgue. En su mayoría este capítulo está basado en [1].

En el Capítulo 3 se presenta una versión integral de la regla de la cadena, llamada la fórmula de Itô. Esta fórmula es muy útil cuando se trata de evaluar integrales de Itô. Este resultado es muy útil cuando se necesita resolver una integral estocástica sin usar la definición presentada en el Capítulo 1. Finalmente se presenta una aplicación de la fórmula de Itô, conocida como la transformación de Lamperti, la cual será de gran utilidad en el Capítulo 8 para un método de estimación en particular

La finalidad del Capítulo 4 es dar a conocer el concepto de ecuación diferencial estocástica. Se hace un estudio sobre las condiciones o supuestos que deben de cumplir los coeficientes de una ecuación diferencial estocástica para que ésta tenga una única solución. Este capítulo está basado en [1] y [2]

En el Capítulo 5 se presentan los procesos que serán estudiados a lo largo de los siguientes capítulos conocidos como procesos de difusión los cuales se abordarán desde dos puntos de vista diferentes. El primero será como solución a una ecuación diferencial estocástica, lo cual es natural, ya que se puede pensar a la dinámica de un activo financiero como una ecuación diferencial que dependa de una función y del tiempo, y otra función que esté perturbada por un proceso estocástico. El segundo enfoque es ver

a un proceso de difusión como a un proceso de Markov, para el cual el estado actual del proceso determina completamente la probabilidad de ocupar los siguientes estados para todos los tiempos futuro. Este capítulo está basado en [2].

En el Capítulo 6 se presentarán los tres proceso de difusión que se usarán para los capítulos de estimación y simulación, así como algunas de sus propiedades, entre ellas, su densidad de transición que como se mencionó anteriormente es la piedra angular para los método de estimación presentados en este trabajo. Este capítulo está basado en [3].

En el Capítulo 7 se presentan aproximaciones numéricas para ecuaciones diferenciales estocásticas, así como sus órdenes de convergencia y las ventajas o desventajas de usar un método u otro. Adicionalmente se hacen las simulaciones para las difusiones presentadas en el capítulo anterior mediante el software R. Este capítulo está completamente basado en [4]

Finalmente en el Capítulo 8 se presentan tres métodos de estimación que son utilizados para las difusiones enunciadas en el Capítulo 6. Se discute cuales son las ventajas y desventajas de usar cada método. Para el método de expansión de la verosimilitud vía polinomios de Hermite se usó [5].

Capítulo 1

Preliminares de procesos estocásticos

En este primer capítulo se mencionarán elementos básicos de la teoría de procesos estocásticos que se usarán a lo largo de este trabajo. En su mayoría se basa en [3].

1.1. Definiciones

Definición 1.1.1. Un proceso estocástico es una colección de variables aleatorias $\mathbf{X} = \{X_t\}_{t \in T}$ definidas en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ parametrizadas en un conjunto T y que toma valores en \mathbb{R} .

El espacio parametral T es usualmente el intervalo $[0, \infty]$.

Notemos que para cada $t \in T$ fijo, tenemos una variable aleatoria

$$X_t(\omega) : \Omega \rightarrow \mathbb{R}, \quad \omega \in \Omega.$$

Por otra parte, fijando $\omega \in \Omega$ podemos considerar la función

$$X_t(\omega) : T \rightarrow \mathbb{R}, \quad t \in T.$$

la cual es llamada una trayectoria o realización de $\mathbf{X} = \{X_t\}_{t \in T}$.

Definición 1.1.2 (Filtración). Una filtración es una familia $(\mathcal{F}_t)_{t \in T}$ de σ -álgebras tales que $\mathcal{F}_s \subset \mathcal{F}_t$ para toda $s \leq t \in T$.

Una filtración puede ser pensada como una estructura de información dinámica. La interpretación es que $(\mathcal{F}_t)_{t \geq 0}$ representa la información disponible al tiempo t .

Definición 1.1.3 (Filtración natural). Se define a la filtración natural de un proceso estocástico $\mathbf{X} = \{X_t\}_{t \geq 0}$ como $\mathcal{F}_t^{\mathbf{X}} \equiv \sigma(X_s, s \leq t)$. Esta notación significa que \mathcal{F}_t es la σ -álgebra más pequeña con respecto a la cual todas las variables $(X_s, s \leq t)$ son medibles.

Definición 1.1.4 (Filtración completa). Si \mathcal{N} es el conjunto de eventos $B \in \mathcal{F}$ tales que $\mathbb{P}(B) = 0$, se define la filtración completa de $\{\mathcal{F}_t^{\mathbf{X}}\}_{t \geq 0}$ mediante la familia de σ -álgebras

$$\mathcal{F}_t = \sigma(\mathcal{F}_t^{\mathbf{X}} \cup \mathcal{N}).$$

Un espacio de probabilidad filtrado se dice que satisface las condiciones usuales, si es completo y continuo por la derecha, es decir, $\mathcal{F}_t = \mathcal{F}_{t+} \equiv \bigcap_{s \geq t} \mathcal{F}_s$ para todo t . Para un conjunto no acotado se define a \mathcal{F}_∞ como la σ -álgebra generada por la unión infinita de \mathcal{F}_t 's, la cual está contenida en \mathcal{F}

$$\mathcal{F}_\infty \equiv \sigma\left(\bigcap_{t > 0} \mathcal{F}_t\right) \subset \mathcal{F}.$$

Definición 1.1.5 (Adaptabilidad). Un proceso estocástico $\mathbf{X} = \{X_t\}_{t \geq 0}$ se dice que es adaptado a una filtración $\{\mathcal{F}_t\}_{t \geq 0}$, si para cada $t \geq 0$, X_t es \mathcal{F}_t -medible, es decir, si para cada $t \geq 0$,

$$\{X_t \leq x\} \equiv \{\omega | X_t(\omega) \leq x\} \in \mathcal{F}_t \text{ para toda } x \in \mathbb{R}.$$

Esto significa que el valor que toma X_t en t depende solamente de la información disponible al tiempo t .

Observación Todo proceso $\mathbf{X} = \{X_t\}_{t \geq 0}$ es adaptado a la filtración $\{\mathcal{F}_t^{\mathbf{X}}\}_{t \geq 0}$.

Definición 1.1.6 (Proceso de Markov). Un proceso estocástico continuo en el tiempo $\mathbf{X} = \{X_t\}_{t \geq 0}$ sobre un espacio de estados numerable E es un proceso de Markov si y sólo si para cualquier $t_0 < t_1 < \dots < t_k < t_{k+1}$ y $C \subset E$, satisface la propiedad de Markov:

$$\mathbb{P}(X_{t_{k+1}} \in C | X_{t_k}, \dots, X_{t_0}) = \mathbb{P}(X_{t_{k+1}} \in C | X_{t_k}).$$

Esto significa que el conocimiento entero de la historia del proceso no contiene más información que el conocimiento de su último valor. Los procesos de Markov describen la evolución temporal de un sistema aleatorio que no tiene memoria.

Definición 1.1.7 (Movimiento browniano). Un proceso estocástico $\mathbf{B} = \{B_t\}_{t \geq 0}$ que satisface las siguientes condiciones se le denomina un movimiento browniano estándar.

(I) $B_0 = 0$ casi donde quiera, es decir,

$$\mathbb{P}[\omega \in \Omega | B_0(\omega) = 0] = 1.$$

(II) Para cualquier conjunto de tiempos $0 \leq t_1 < t_2 < \dots < t_n$, los incrementos

$$B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}},$$

son estocasticamente independientes.

(III) Para cualquier par de tiempos t y s con $0 \leq s < t$, $B_t - B_s \sim N(0, t - s)$.

Definición 1.1.8 (Martingala). Un proceso estocástico $\mathbf{M} = \{M_t\}_{t \geq 0}$ definido sobre un espacio de probabilidad, $(\Omega, \mathcal{F}, \mathbb{P})$, se denomina martingala con respecto a una filtración $\{\mathcal{F}_t\}_{t \in T}$ (y con respecto a \mathbb{P}) si se cumple que

(I) \mathbf{M} es un proceso adaptado a la filtración $\{\mathcal{F}_t\}_{t \in T}$, es decir, para cada $t \geq 0$, M_t es \mathcal{F}_t -medible.

(II) $\mathbb{E}[|M_t|] < \infty, \quad \forall t > 0$.

(III) $\mathbb{E}[M_s | \mathcal{F}_t] = M_t, \quad \forall s \geq t$.

Si (II) no se cumple, se dice que $\mathbf{M} = \{M_t\}$ es una martingala local.

Ejemplo 1.1.9. El movimiento browniano sobre un espacio fijo de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ equipado con su filtración aumentada es una martingala.

(I) Se cumple debido a que todo proceso estocástico es adaptado a su filtración natural.

$$\begin{aligned} \text{(II)} \quad \mathbb{E}[B_t | \mathcal{F}_s] &= \mathbb{E}[B_t - B_s + B_s | \mathcal{F}_s] \\ &= \mathbb{E}[(B_t - B_s) + B_s | \mathcal{F}_s] \\ &= \mathbb{E}[B_t - B_s | \mathcal{F}_s] + B_s \\ &= 0 + B_s. \end{aligned}$$

Aquí usamos que $\mathbb{E}[(B_s - B_t) | \mathcal{F}_t] = \mathbb{E}[B_s - B_t] = 0$ pues $B_t - B_s$ es independiente de \mathcal{F}_t y que $\mathbb{E}[B_t | \mathcal{F}_t] = B_t$ debido a que B_t es \mathcal{F}_t -medible.

$$\begin{aligned} \text{(III)} \quad \mathbb{E}[B_t] &= \int_{-\infty}^{\infty} |\omega| d\mathbb{P}(w) \\ &= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} |\omega| e^{-\frac{1}{2t}\omega^2} d\omega \\ &= \frac{1}{\sqrt{2\pi t}} \int_0^{\infty} 2\omega e^{-\frac{1}{2t}\omega^2} d\omega \\ &= \frac{1}{\sqrt{2\pi t}} \int_0^{\infty} e^{-\frac{1}{2t}y} dy \quad (\text{haciendo } y = \omega^2) \\ &= \sqrt{\frac{2t}{\pi}} < \infty. \end{aligned}$$

Capítulo 2

Cálculo de Itô

En este capítulo se busca dar una breve introducción al Cálculo de Itô. En la primera sección se construye la integral de Itô y al final de éste se resuelve una integral usando la definición. En la segunda sección se presenta la fórmula de Itô, una herramienta de bastante utilidad que se usará en capítulos posteriores, la cual permite llegar a evaluar integrales de Itô de forma sencilla. Este capítulo está basado en su totalidad en [1].

2.1. Construcción de la integral de Itô

Consideremos el modelo de crecimiento poblacional dado por:

$$\frac{dN_t}{dt} = a(t)N_t; \quad N(0) = N_0 \text{ (constante)}, \quad (2.1)$$

donde N_t es el tamaño de la población al tiempo t , y $a(t)$ es la tasa de crecimiento relativo al tiempo t . Podría suceder que $a(t)$ no se conozca completamente debido a factores aleatorios. Así, se tiene que

$$a(t) = r(t) + \text{“ruido”},$$

donde no se sabe exactamente el comportamiento del “ruido”, solo su distribución de probabilidad. La función $r(t)$ se asume no aleatoria. ¿Cómo se podrá resolver la ecuación (2.1)? o más general, ¿cómo se podrían resolver ecuaciones de la forma

$$\frac{dX_t}{dt} = b(t, X_t) + \sigma(t, X_t) \cdot \text{“ruido”}, \quad (2.2)$$

donde b y σ son funciones dadas. Se podría pensar que algún proceso estocástico $\mathbf{W} = \{W_t\}_{t \geq 0}$ como candidato para modelar el “ruido”, de modo que (2.2) se puede escribir

como:

$$\frac{dX_t}{dt} = b(t, X_t) + \sigma(t, X_t) \cdot W_t. \quad (2.3)$$

Se asume que $\mathbf{W} = \{W_t\}_{t \geq 0}$ tiene, al menos que satisfacer las siguientes propiedades:

- a) Si $t_1 \neq t_2$ entonces W_{t_1} y W_{t_2} son independientes.
- b) El proceso $\mathbf{W} = \{W_t\}_{t \geq 0}$ es estacionario, es decir, la distribución conjunta de $(W_{t_1+t}, \dots, W_{t_k+t})$ no dependen de t .
- c) $\mathbb{E}[W_t] = 0$ para toda $t > 0$.

Sin embargo, se puede demostrar que no existe ningún proceso $\mathbf{W} = \{W_t\}_{t \geq 0}$ que satisfaga las propiedades anteriores tal que $\mathbf{W} = \{W_t\}_{t \geq 0}$ tenga trayectorias continuas. No obstante es posible representar a $\mathbf{W} = \{W_t\}_{t \geq 0}$ como un proceso estocástico generalizado conocido como el proceso de ruido blanco.

A continuación se tratará de escribir a la ecuación (2.3) de tal forma que se pueda reemplazar $\mathbf{W} = \{W_t\}_{t \geq 0}$ por un proceso estocástico adecuado. Sean $0 = t_0 < t_1 < \dots < t_m = t$ una partición de $[0, t]$ y considérese una versión discreta de la ecuación (2.3)

$$X_{k+1} - X_k = b(t_k, X_k)\Delta t_k + \sigma(t_k, X_k)W_k\Delta t_k, \quad (2.4)$$

donde

$$X_k = X_{t_k}, \quad W_k = W_{t_k}, \quad \Delta_k = t_{k+1} - t_k.$$

Reemplazando $W_k\Delta_k t_k$ por $\Delta V_k = V_{t_{k+1}} - V_{t_k}$, donde $\{V_t\}_{t \geq 0}$ es algún proceso adecuado. Los supuestos (a), (b) y (c) sugieren que $\{V_t\}_{t \geq 0}$ debe tener incrementos estacionarios independientes con media 0. Se puede demostrar que el único proceso que satisface estas condiciones con trayectorias continuas es el movimiento browniano $\mathbf{B} = \{B_t\}_{t \geq 0}$. De la expresión (2.4) y sustituyendo V_t por B_t , se tiene que:

$$X_k = X_0 + \sum_{j=0}^{k-1} b(t_j, X_j)\Delta t_j + \sum_{j=0}^{k-1} \sigma(t_j, X_j)\Delta B_j. \quad (2.5)$$

Si el límite del lado derecho de (2.5) existiera en algún sentido cuando $\Delta t_j \rightarrow 0$, entonces análogamente se podría aplicar la notación de integración usual, es decir,

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s. \quad (2.6)$$

Se tendría que la ecuación (2.2) significa que $\mathbf{X}=\{X_t\}_{t \geq 0}$ es un proceso estocástico que satisface la ecuación (2.6).

A continuación se demostrará la existencia, en cierto sentido, de:

$$\int_0^t f(s, X_s) dB_s \tag{2.7}$$

donde $\mathbf{B} = \{B_t\}_{t \geq 0}$, es un movimiento browniano, que inicia en el 0, para una clase grande de funciones $f : [0, \infty] \times \Omega \rightarrow \mathbb{R}$.

Análogamente a la construcción de la integral de Lebesgue se empezará la definición de la expresión (2.7) para una clase simple de funciones y después se extenderá mediante algún método de aproximación.

Supongamos $0 \leq S < T$ y $f(t, \omega)$ dados y para empezar supongamos que f tiene la forma:

$$f(t, \omega) = \sum_{j \geq 0} e_j(\omega) \cdot \mathbb{I}_{[\frac{j}{2^n}, \frac{j+1}{2^n}]}(t); \quad n \in \mathbb{N}, \tag{2.8}$$

donde \mathbb{I} denota la función indicadora y $e_j(\cdot)$ es \mathcal{F}_j -medible.

Para tales funcione se define

$$\int_S^T f(t, \omega) dB_t = \sum_{j \geq 0} e_j([B_{t_{j+1}} - B_{t_j}]), \tag{2.9}$$

donde

$$t_k = t_k^{(n)} = \begin{cases} k \cdot 2^{-n} & \text{si } S \leq k \cdot 2^{-n} \leq T \\ S & \text{si } k \cdot 2^{-n} < S \\ T & \text{si } k \cdot 2^{-n} > T. \end{cases} \tag{2.10}$$

Como se verá en los siguientes ejemplos, si no se hacen otros supuestos sobre las funciones $e_j(\omega)$, esto conducirá a dificultades.

Ejemplo 2.1.1. Supongamos que se definen $f_1(t, \omega)$ y $f_2(t, \omega)$ como sigue:

$$f_1(t, \omega) = \sum_{j \geq 0} B_{\frac{j}{2^n}} \cdot \mathbb{I}_{[\frac{j}{2^n}, \frac{j+1}{2^n}]}(t)$$

$$f_2(t, \omega) = \sum_{j \geq 0} B_{\frac{j+1}{2^n}} \cdot \mathbb{I}_{[\frac{j}{2^n}, \frac{j+1}{2^n}]}(t)$$

Entonces, debido a que el proceso $B_t = \{B_t\}_{t \geq 0}$ tiene incrementos independientes, se tiene que

$$\mathbb{E} \left[\int_0^T f_1(t, \omega) dB_t \right] = \sum_{j \geq 0} \mathbb{E}[B_{t_j}(B_{t_{j+1}} - B_{t_j})] = 0. \quad (2.11)$$

Por otro lado

$$\begin{aligned} \mathbb{E} \left[\int_0^T f_2(t, \omega) dB_t \right] &= \sum_{j \geq 0} \mathbb{E}[B_{t_{j+1}}(B_{t_{j+1}} - B_{t_j})] \\ &= \sum_{j \geq 0} \mathbb{E}[(B_{t_{j+1}} - B_{t_j})^2] = T. \end{aligned}$$

puesto que $\mathbb{E}[(B_t - B_s)^2] = n(t - s)$.

Entonces, a pesar de que f_1 y f_2 parecen ser aproximaciones muy razonables a

$$f(t, \omega) = B_t,$$

sus integrales de acuerdo a la igualdad (2.9), no serían iguales, sin importar que tan grande n sea escogida. Esto sugiere que las variaciones de las trayectorias de $B_t = \{B_t\}_{t \geq 0}$ son demasiado grandes para definir la integral

$$\int_S^T f(t, \omega) dB_t$$

de la misma forma que una integral de Riemann-Stieltjes.

En general se aproxima una función dada $f(t, \omega)$ por

$$\sum_j f(t_j^*, \omega) \cdot \mathbb{I}_{[t_j, t_{j+1})}(t),$$

donde los puntos t_j^* pertenecen a los intervalos $[t_j, t_{j+1}]$, y entonces se define a

$$\int_S^T f(t, \omega) dB_t$$

como el límite de $\sum_j f(t_j^*, \omega)[B_{t_{j+1}} - B_{t_j}]$ cuando n converge a ∞ . Sin embargo, a diferencia de la integral de Riemann-Stieltjes, la elección de los puntos t_j^* sí importa. Las siguientes dos elecciones de t_j^* han resultado ser las más útiles.

a) Eligiendo $t_j^* = t_j$, nos conduce a la integral de Itô denotada como sigue

$$\int_S^T f(t, \omega) dB_t,$$

- b) Eligiendo $t_j^* = \frac{t_j + t_{j+1}}{2}$, nos conduce a la integral de Stratonovich denotada como sigue

$$\int_S^T f(t, \omega) \circ dB_t.$$

A continuación se explicará porque se usa la elección de Itô $t_j^* = t_j$. El procedimiento de aproximación sólo funcionará si f tiene la propiedad de que cada una de las funciones $\omega \rightarrow f(t_j, \omega)$ sólo depende del comportamiento de $\{B_s\}_{s \geq 0}$ hasta el tiempo t_j . Esto lleva al siguiente concepto importante.

Definición 2.1.2. Sea $\mathbf{V} = \mathcal{V}(S, T)$ la clase de funciones

$$f(t, \omega) : [0, \infty) \times \Omega \rightarrow \mathbb{R}$$

tales que

(I) $(t, \omega) \Rightarrow f(t, \omega)$ es $\mathcal{B} \times \mathcal{F}$ -medibles, donde \mathcal{B} denota la σ -álgebra sobre $[0, \infty)$.

(II) $f(t, \omega)$ es \mathcal{F}_t -adaptado

(III) $\mathbb{E}[\int_S^T f(t, \omega)^2 dt] < \infty$.

2.1.1. Integral de Itô

Para funciones $f \in \mathbf{V}$ se mostrará como definir la integral de Itô

$$\int_S^T f(t, \omega) dB_t,$$

donde $B_t = \{B_t\}_{t \geq 0}$ es un movimiento browniano. La idea es análoga a la construcción de la integral de Lebesgue en el sentido de que se define la integral de Itô para una cierta clase de funciones elementales ϕ , luego se verá que para cada $f \in \mathbf{V}$, f puede ser aproximada (en un sentido apropiado) por las ϕ 's y se define

$$\int f dB_t := \lim_{\phi \rightarrow f} \int \phi dB_t.$$

Definición 2.1.3. Una función $\phi \in \mathbf{V}$ se denomina elemental si tiene la siguiente forma

$$\phi(t, \omega) = \sum_j e_j(\omega) \cdot \mathbb{I}_{[t_j, t_{j+1})}(t),$$

donde $e_j : \Omega \rightarrow \mathbb{R}$, para todo $j \in \mathbb{N}$.

Observación 2.1.4. Debido a que $\phi \in \mathbf{V}$, entonces cada función e_j debe ser \mathcal{F}_{t_j} -medible. Así, en el Ejemplo 2.1.1, la función f_1 es elemental, mientras f_2 no lo es.

Para funciones elementales $\phi(t, \omega)$ se define su integral respecto al movimiento browniano como:

$$\int_S^T \phi(t, \omega) dB_t = \sum_{j \geq 0} e_j(\omega) [B_{t_{j+1}} - B_{t_j}],$$

donde $n \in \mathbb{N}$ y $S = t_0 < t_1 < \dots < t_{n-1} < t_n = T$ es una partición del intervalo $[S, T]$.

Observación 2.1.5. Cada función $e_j(\omega)$ es independientes de $[B_{t_{j+1}} - B_{t_j}]$ como resultado de los incrementos independientes del movimiento browniano $[B_{t_{j+1}} - B_{t_j}]$ y por la observación anterior.

A continuación se presentarán algunas propiedades importantes sobre la integral de Itô.

Lema 2.1.6 (Isometría de Itô). Si $\phi(t, \omega)$ es acotada y elemental, entonces

$$\mathbb{E} \left[\left(\int_S^T \phi(t, \omega) dB_t \right)^2 \right] = \mathbb{E} \left[\int_S^T \phi(t, \omega)^2 dt \right]. \quad (2.12)$$

Demostración. Definiendo $\Delta B_j = B_{t_{j+1}} - B_{t_j}$. Entonces se tiene que

$$\mathbb{E}[e_i e_j \Delta B_i \Delta B_j] = \begin{cases} 0 & \text{si } i \neq j \\ \mathbb{E}[e_j^2] \cdot (t_{j+1} - t_j) & \text{si } i = j \end{cases}$$

debido a que $e_i e_j \Delta B_i$ y ΔB_j son independientes si $i < j$.

Por lo tanto

$$\begin{aligned} \mathbb{E} \left[\left(\int_S^T \phi dB_t \right)^2 \right] &= \sum_{i,j} \mathbb{E}[e_i e_j \Delta B_i \Delta B_j] = \sum_j \mathbb{E}[e_j^2] \cdot (t_{j+1} - t_j) \\ &= \mathbb{E} \left[\int_S^T \phi^2 dt \right]. \end{aligned}$$

□

Observación 2.1.7. Si $\phi = \phi(t, \omega) \in \mathbf{V}$, entonces

$$\mathbb{E} \left[\int_S^T \phi(t, \omega) dB_t \right] = \mathbb{E} \left[\sum_{j=0}^{n-1} e_j(\omega) [B_{t_{j+1}} - B_{t_j}] \right] = 0,$$

puesto que $e_j(\omega)$ es independiente de $[B_{t_{j+1}} - B_{t_j}](\omega)$ y $\mathbb{E}[B_{t_{j+1}} - B_{t_j}] = 0$.

La idea es usar el Lema 2.1.6 para extender la definición de integral para funciones elementales $\phi(t, \omega)$ a funciones en \mathbf{V} .

Paso 1 Sea $g \in \mathbf{V}$ acotada y $g(\cdot, \omega)$ continua para cada ω . Entonces, existen funciones elementales $\{\phi_n\}_{n \geq 1} \in \mathbf{V}$ tales que

$$\mathbb{E} \left[\int_S^T (g - \phi_n)^2 dt \right] \rightarrow 0 \quad \text{cuando } n \rightarrow \infty.$$

Demostración. Definiendo $\phi_n(t, \omega) = \sum_j g(t_j, \omega) \cdot \mathbb{I}_{[t_j, t_{j+1}]}(t)$. Entonces ϕ_n es elemental pues $g \in \mathbf{V}$ y

$$\int_S^T (g - \phi_n)^2 \rightarrow 0 \quad \text{cuando } n \rightarrow \infty, \text{ para cada } \omega.$$

cuando $g(\cdot, \omega)$ es continua para cada ω . Por lo tanto $\mathbb{E} \left[\int_S^T (g - \phi_n)^2 dt \right] \rightarrow 0$ cuando $n \rightarrow \infty$, por el teorema de convergencia acotada. \square

Paso 2 Sea $h \in \mathbf{V}$ acotada. Entonces existen funciones acotadas $g_n \in \mathbf{V}$ tales que $g_n(\cdot, \omega)$ es continua para todo ω, n , y

$$\mathbb{E} \left[\int_S^T (h - g_n)^2 dt \right] \rightarrow 0.$$

Demostración. Como h es acotada, supongase $|h(t, \omega)| \leq M$ para todo (t, ω) . Para cada n sea ψ_n una función no negativa, continua en \mathbb{R} tal que se satisfacen las siguientes condiciones

- (I) $\psi_n(x) = 0$ para $x \leq -\frac{1}{n}$ y $x \geq 0$.
- (II) $\int_{-\infty}^{\infty} \psi_n(x) = 1$.

definiendo

$$g_n(t, \omega) = \int_0^t \psi_n(s-t) h(s, \omega) ds.$$

Entonces $g_n(\cdot, \omega)$ es continua para cada ω y $|g_n(t, \omega)| \leq M$. Puesto que $h \in \mathbf{V}$ se tiene que $g_n(t, \cdot)$ es \mathcal{F}_t -medible para toda t . Más aún,

$$\int_S^T (g_n(s, \omega) - h(s, \omega))^2 \rightarrow 0 \quad \text{cuando } n \rightarrow \infty, \text{ para cada } \omega,$$

puesto que $\{\psi_n\}_n$ constituye una identidad aproximada. Así entonces por el teorema de convergencia acotada

$$\mathbb{E} \left[\int_S^T (h(t, \omega) - g_n(t, \omega))^2 dt \right] \rightarrow 0.$$

□

Paso 3 Sea $f \in \mathbf{V}$. Entonces existe una sucesión $\{h_n\} \subset \mathbf{V}$ tal que h_n es acotada para cada n y

$$\mathbb{E} \left[\int_S^T (f - h_n)^2 dt \right] \rightarrow 0 \quad \text{cuando } n \rightarrow \infty.$$

Demostración. Sea

$$h_n(t, \omega) = \begin{cases} -n & \text{si } f(t, \omega) < -n \\ f(t, \omega) & \text{si } -n \leq f(t, \omega) \leq n \\ n & \text{si } f(t, \omega) > n \end{cases} \quad (2.13)$$

La demostración se sigue por el teorema de convergencia dominada. □

Ahora se puede completar la definición de la integral de Itô.

Definición 2.1.8 (Integral de Itô). Sea $f \in \mathbf{V}$. Entonces la integral de Itô de f (desde S a T) se define como

$$\int_S^T f(t, \omega) dB_t := \lim_{n \rightarrow \infty} \int_S^T \phi_n(t, \omega) dB_t$$

donde $\{\phi\}_{n \geq 1} \in \mathbf{V}$ es una sucesión de funciones elementales tales que

$$\mathbb{E} \left[\int_S^T |f - \phi_n|^2 dt \right] \rightarrow 0. \quad (2.14)$$

El límite existe como un elemento de $L^2(P)$, puesto que

$$\left\{ \int_S^T \phi_n(t, \omega) dB_t \right\}$$

forma una sucesión de Cauchy en $L^2(P)$, por la Isometría de Itô.

A continuación se presenta un ejemplo de esta integral.

Ejemplo 2.1.9. Suponiendo $B_0 = 0$. Entonces

$$\int_0^t B_s dB_s = \frac{1}{2} B_t^2 - \frac{1}{2} t$$

Demostración. Primero se verifica que la condición del **Paso 3** se satisface.

Se define a $\phi_n(s, \omega) := \sum_{j=0}^{n-1} B_{t_j}(\omega) \mathbb{I}_{[t_j, t_{j+1})}(s)$

Entonces

$$\begin{aligned} \mathbb{E} \left[\int_0^t (\phi_n - B_{t_s})^2 ds \right] &= \mathbb{E} \left[\sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} (B_{t_j} - B_{t_s})^2 ds \right] \\ &= \sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} \mathbb{E} \left[(B_{t_s} - B_{t_j})^2 \right] ds \\ &= \sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} (s - t_j) ds \quad \left(\mathbb{E} \left[(B_{t_s} - B_{t_j})^2 \right] = \text{Var}(B_{t_s} - B_{t_j}) \right) \\ &= \sum_{j=0}^{n-1} \left[\frac{1}{2} s^2 - s \cdot t_j \right]_{s=t_j}^{s=t_{j+1}} \\ &= \sum_{j=0}^{n-1} \left(\frac{1}{2} t_{j+1}^2 - t_j \cdot t_{j+1} - \frac{1}{2} t_j^2 + t_j^2 \right) \\ &= \sum_{j=0}^{n-1} \frac{1}{2} (t_{j+1} - t_j)^2 \rightarrow 0 \quad \text{cuando } \Delta t_j := (t_{j+1} - t_j) \rightarrow 0. \end{aligned}$$

Por el **Paso 3** se tiene que

$$\int_0^t B_s dB_s = \lim_{\Delta t_j \rightarrow 0} \int_0^t \phi_n dB_s = \lim_{\Delta t_j \rightarrow 0} \sum_{j=0}^{n-1} B_j \Delta B_j.$$

Además

$$\Delta(B_{t_j}^2) = B_{t_{j+1}}^2 - B_{t_j}^2 = (B_{t_{j+1}} - B_{t_j})^2 + 2B_{t_j}(B_{t_{j+1}} - B_{t_j}) = (\Delta B_{t_j})^2 + 2B_{t_j} \Delta B_{t_j}.$$

Por lo tanto, como $B_0 = 0$, se tiene que

$$B_t^2 = \sum_{j=0}^{n-1} \Delta(B_{t_j}^2) = \sum_{j=0}^{n-1} (\Delta B_{t_j})^2 + 2 \sum_{j=0}^{n-1} B_{t_j} \Delta B_j.$$

Finalmente

$$\begin{aligned} \sum_{j=0}^{n-1} B_{t_j} \Delta B_j &= \frac{1}{2} \sum_{j=0}^{n-1} \Delta(B_{t_j}^2) - \frac{1}{2} \sum_{j=0}^{n-1} (\Delta B_{t_j})^2 \\ &= \frac{1}{2} B_t^2 - \frac{1}{2} \sum_{j=0}^{n-1} (\Delta B_{t_j})^2 \\ &= \frac{1}{2} B_t^2 - \frac{1}{2} t. \quad \left(\sum_{j=0}^{n-1} (\Delta B_{t_j})^2 \rightarrow t \text{ en } L^2 \text{ cuando } \Delta t_j \rightarrow 0 \right). \end{aligned}$$

□

Observación 2.1.10. El término $-\frac{1}{2}t$ indica que la integral estocástica de Itô no tiene un comportamiento ordinario como las otras integrales.

2.2. Fórmula de Itô

En el ejemplo 2.1.9 se analizó una forma de resolver una integral de Itô a través de su definición, la cuál no es útil cuando se quiere evaluar una integral dada. De manera similar al caso en el que la definición de integrales ordinarias de Riemann no es muy útil para resolver este tipo de integrales, donde en lugar de usar la definición, se usa el teorema fundamental del cálculo junto con la regla de la cadena para cálculos explícitos. Sin embargo en el contexto de integrales de Itô, no se tiene una teoría de diferenciación, sólo se tiene una teoría de integración. Sin embargo, resulta que es posible establecer una versión integral de la regla de la cadena, llamada la fórmula de Itô. Esta fórmula es muy útil cuando se trata de evaluar integrales de Itô.

Del ejemplo 2.1.9

$$\int_0^t B_s dB_s = \frac{1}{2} B_t^2 - \frac{1}{2},$$

se mostró que la imagen de la integral de Itô $B_t = \int_0^t dB_s$ por el mapeo $g(x) = \frac{1}{2}x^2$ no resulta ser de nuevo una integral de Itô de la forma

$$\int_0^t f(s, \omega) dB_s,$$

si no, una combinación de integrales respecto a dB_s y ds

$$\frac{1}{2} B_t^2 = \int_0^t \frac{1}{2} ds + \int_0^t B_s dB_s.$$

Definición 2.2.1. Sea $\mathcal{W} = \mathcal{W}_{\mathcal{H}}(S, T)$ la clase de procesos $f(t, \omega) \in \mathbb{R}$ que satisface

(I) $(t, \omega) \Rightarrow f(t, \omega)$ es $\mathcal{B} \times \mathcal{F}$ -medibles, donde \mathcal{B} denota la σ -álgebra sobre $[0, \infty)$.

(II) Existe una familia creciente de σ -álgebras $\{\mathcal{H}_t; t \geq 0$ tales que

a) El proceso $\mathbf{B} = \{B_t\}_{t \geq 0}$ es martingala con respecto a \mathcal{H}_t .

b) f_t es \mathcal{H}_t -adaptado.

(III) $\mathbb{P} \left[\int_S^T f(s, \omega)^2 ds < \infty \right] = 1$.

Definición 2.2.2 (Procesos de Itô). Sea $B = \{B_t\}_{t \geq 0}$ un movimiento browniano sobre $(\Omega, \mathcal{F}, \mathbb{P})$. Un proceso de Itô (o integral estocástica) es un proceso estocástico $\mathbf{X} = \{X_t\}_{t \geq 0}$ sobre $(\Omega, \mathcal{F}, \mathbb{P})$ de la forma

$$X_t = X_0 + \int_0^t u(s, \omega) ds + \int_0^t v(s, \omega) dB_s, \quad (2.15)$$

donde $v \in W_{\mathcal{H}}$, tal que

$$\mathbb{P} \left[\int_0^t v(s, \omega)^2 ds < \infty \right] = 1 \quad \text{para todo } t \geq 0.$$

Se supone también que u es \mathcal{H}_t -adaptado (donde \mathcal{H}_t es como en (II) de la definición 2.2.1) y

$$\mathbb{P} \left[\int_0^t |u(s, \omega)| ds < \infty \right] = 1 \quad \text{para todo } t \geq 0.$$

Si $\mathbf{X} = \{X_t\}_{t \geq 0}$ es un proceso solución de la forma (2.15), la ecuación (2.15) usualmente se escribe en forma corta como

$$dX_t = u dt + v dB_t.$$

por ejemplo, la ecuación (2.2) se puede escribir como:

$$d\left(\frac{1}{2}B_t^2\right) = \frac{1}{2}dt + B_t dB_t.$$

A continuación se presentará el resultado principal de este capítulo.

Teorema 2.2.3 (Fórmula de Itô). Sea $\mathbf{X} = \{X_t\}_{t \geq 0}$ un proceso de Itô dado por

$$dX_t = u dt + v dB_t.$$

Sea $g(t, x) \in C^2([0, \infty) \times \mathbb{R})$. Entonces

$$Y_t = g(t, X_t)$$

es de nuevo un proceso de Itô y

$$dY_t = \frac{\partial g}{\partial t}(t, X_t)dt + \frac{\partial g}{\partial x}(t, X_t)dX_t + \frac{1}{2} \frac{\partial^2 g}{\partial x^2}(t, X_t) \cdot (dX_t)^2, \quad (2.16)$$

donde $(dX_t)^2 = (dX_t) \cdot (dX_t)$ se calcula de acuerdo a las reglas de variación y covariación cuadráticas.

$$dt \cdot dt = dt \cdot dB_t = dB_t \cdot dt = 0, \quad dB_t \cdot dB_t = dt. \quad (2.17)$$

El siguiente resultado es la versión del calculo de Itô para integrar una función por partes.

Teorema 2.2.4 (Integración por partes). Sea $f(s, \omega) = f(s)$ (sólo depende de s) y que f es continua y de variación acotada en $[0, t]$.

Entonces

$$\int_0^t f(s)dB_s = f(t)B_t - \int_0^t B_sdf_s. \quad (2.18)$$

Observación 2.2.5. Es crucial para el resultado anterior que f no dependa de ω .

A continuación se verán dos ejemplos del uso de los teoremas mencionado anteriormente

Ejemplo 2.2.6. Regresando al último ejemplo de la sección anterior

$$I = \int_0^t B_sdB_s.$$

Se tiene que $\mathbf{X} = \{X_t\}_{\geq 0} = B_t = \{B_t\}_{\geq 0}$ y $g(t, x) = \frac{1}{2}x^2$.

Entonces

$$Y_t = g(t, B_t) = \frac{1}{2}B_t^2.$$

Entonces por la fórmula de Itô,

$$\begin{aligned} dY_t &= \frac{\partial g}{\partial t}dt + \frac{\partial g}{\partial x}dB_t + \frac{1}{2} \frac{\partial^2 g}{\partial x^2}(dB_t)^2 \\ &= B_tdB_t + \frac{1}{2}(dB_t)^2 \\ &= B_tdB_t + \frac{1}{2}dt. \end{aligned}$$

Por lo tanto

$$d\left(\frac{1}{2}B_t^2\right) = B_t dB_t + \frac{1}{2}dt.$$

O bien,

$$\frac{1}{2}B_t^2 = \int_0^t B_s dB_s + \frac{1}{2}t.$$

Ejemplo 2.2.7.

$$I = \int_0^t s dB_s,$$

debido a que el término tB_t debe aparecer, se tiene que $g(t, x) = tx$ y

$$Y_t = g(t, B_t) = tB_t.$$

Por la fórmula de Itô

$$dY_t = B_t dt + t dB_t + 0,$$

es decir,

$$d(tB_t) = B_t dt + t dB_t,$$

o bien,

$$tB_t = \int_0^t B_s ds + \int_0^t s dB_s.$$

Lo cual se obtiene inmediatamente usando el teorema de Integración por partes, notando que g sólo depende de t .

El siguiente resultado, aplicación del Lema de Itô, será muy útil en el capítulo de estimación cuando se necesite hacer el coeficiente de difusión unitario $\sigma = 1$ en una ecuación diferencial estocástica.

2.2.1. Transformación de Lamperti

Existe una aplicación particular del Lema de Itô que es de interés en métodos de simulación y estimación para procesos de difusión. Supongamos que tenemos la ecuación diferencial estocástica

$$dX_t = b(t, X_t)dt + \sigma(X_t)dB_t,$$

donde el coeficiente de difusión depende sólo sobre el estado de la variable. Tal ecuación diferencial estocástica puede ser siempre transformada en otra ecuación diferencial estocástica con coeficiente de difusión unitario aplicando la **transformación de Lamperti**,

$$Y_t = F(\mathbf{X}) = \int_z^{X_t} \frac{1}{\sigma(u)} du. \tag{2.19}$$

z es cualquier valor arbitrario en el espacio de estados de $\mathbf{X} = \{X_t\}_{\geq 0}$. Aún más, el proceso $\mathbf{Y} = \{Y_t\}_{\geq 0}$ resuelve la ecuación diferencial estocástica

$$dY_t = b_y(t, Y_t)dt + dB_t,$$

donde

$$b_y(t, y) = \frac{b(t, F^{-1}(y))}{\sigma(F^{-1}(y))} - \frac{1}{2}\sigma_x(F^{-1}(y)),$$

la cuál también se puede escribir como

$$dY_t = \left(\frac{b(t, X_t)}{\sigma(X_t)} - \frac{1}{2}\sigma_x(X_t) \right) dt + dB_t. \quad (2.20)$$

Para obtener este resultados, se debe aplica la fórmula de Itô a $f(t, x) = \int_z^x \frac{1}{\sigma(u)} du$, donde

$$f_t(t, x) = 0 \quad f_x(t, x) = \frac{1}{\sigma(x)}, \quad f_{xx}(t, x) = \frac{-\sigma_x(x)}{\sigma^2(x)}.$$

Entonces,

$$\begin{aligned} df(t, x) &= 0 \cdot dt + f_x(t, X_t)dB_t + \frac{1}{2}f_{xx}(t, X_t)(dX_t)^2 \\ &= \frac{b(t, X_t)dt + \sigma(X_t)dB_t}{\sigma(X_t)} - \frac{1}{2} \frac{\sigma_x(t, X_t)}{\sigma^2(X_t)} (b(t, X_t)dt + \sigma(X_t)dB_t)^2 \\ &= \frac{b(t, X_t)}{\sigma(X_t)} dt + dB_t - \frac{1}{2} \frac{\sigma_x(t, X_t)}{\sigma^2(X_t)} \times (b(t, X_t)(dt)^2 + 2 \cdot b(t, X_t)\sigma(X_t)dt \cdot dB_t \\ &\quad + \sigma^2(x)(dB_t)^2). \end{aligned}$$

Capítulo 3

Ecuaciones diferenciales estocásticas

3.1. Introducción

En este capítulo se buscará responder preguntas acerca de la existencia y la unicidad de soluciones para ecuaciones diferenciales estocásticas, así como presentar algunas propiedades. Este estudio es realmente acerca de procesos de difusión. El término difusión se atribuye a procesos de Markov con trayectorias muestrales continuas y que puede ser caracterizado a través de su generador infinitesimal.

Regresando a las posibles soluciones $\mathbf{X} = \{X_t\}_{t \geq 0}$ de la ecuación diferencia estocástica

$$\frac{dX_t}{dt} = b(t, X_t) + \sigma(t, X_t)W_t, \quad b(t, x), \sigma(t, x) \in \mathbb{R}. \quad (3.1)$$

donde \mathbf{W} es un “ruido blanco”. Como se menciona en el Capítulo 2 la interpretación de la ecuación (3.1) es que $\mathbf{X} = \{X_t\}_{t \geq 0}$ satisface la ecuación integral estocástica

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s, \quad (3.2)$$

o en su forma diferencial

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t. \quad (3.3)$$

Por lo tanto, para pasar de la expresión (3.1) a la expresión (3.3), se puede pensar al ruido blanco como la derivada de un movimiento browniano, es decir, reemplazando a W_t

por $\frac{dB_t}{dt}$ en la ecuación (3.1) y multiplicando por dt . Análogamente al trato que se da en ecuaciones diferenciales ordinarias, es natural preguntarse: ¿Cómo se pueden resolver estas ecuaciones? ó ¿si existen teoremas de existencia y unicidad para tales ecuaciones? y ¿qué propiedades pueden tener estas soluciones?.

Como se pudo observar en la sección 2.2, la fórmula de Itô es la clave para resolver muchas ecuaciones diferenciales estocásticas. A continuación se mostrará este método para un ejemplo, cuya solución es un proceso estocástico muy usado en el mundo financiero.

Ejemplo 3.1.1. Regresando al modelo de crecimiento poblacional

$$\frac{dN_t}{dt} = a_t N_t, \quad N_0 \text{ dado,}$$

donde $a_t = r_t + \alpha W_t$, W_t = ruido blanco, α constante.

Supongamos que $r_t = r$ (constante para cualquier t). Por la interpretación de Itô en su forma diferencial, la ecuación del modelo poblacional es equivalente a

$$dN_t = rN_t dt + \alpha N_t dB_t,$$

ó

$$\frac{dN_t}{N_t} = r dt + \alpha dB_t.$$

Por lo tanto

$$\int_0^t \frac{dN_s}{N_s} = rt + \alpha B_t \quad (B_0 = 0).$$

Para evaluar esta integral, usamos la fórmula de Itô para la función

$$g(t, x) = \ln x; \quad x > 0.$$

Calculando las respectivas derivadas parciales, se obtiene que

$$\begin{aligned} d(\ln N_t) &= \frac{1}{N_t} \cdot dN_t + \frac{1}{2} \left(-\frac{1}{N_t^2} \right) (dN_t)^2 \\ &= \frac{dN_t}{N_t} - \frac{1}{2N_t^2} \cdot \alpha^2 N_t^2 dt \\ &= \frac{dN_t}{N_t} - \frac{1}{2} \alpha^2 dt. \end{aligned}$$

Así,

$$\frac{dN_t}{N_t} = d(\ln N_t) + \frac{1}{2} \alpha^2 dt.$$

Se concluye que

$$\ln \frac{N_t}{N_0} = \left(r - \frac{1}{2}\alpha^2\right)t + \alpha B_t$$

ó

$$N_t = N_0 \exp\left(\left(t - \frac{1}{2}\alpha^2\right)t + \alpha B_t\right). \quad (3.4)$$

Tal proceso solución es llamado el movimiento browniano geométrico.

3.2. Un resultado de existencia y unicidad para ecuaciones diferenciales estocásticas

Regresando a la pregunta acerca de existencia y unicidad para soluciones para ecuaciones diferenciales estocásticas, tomando en cuenta los siguientes supuestos, es posible dar un teorema de existencia y unicidad para ecuaciones diferenciales estocásticas.

Si el coeficiente de difusión (matriz de dispersión) $\sigma(t, x)$ es igual a cero, entonces la ecuación integral (3.2) se reduce a la ecuación integral ordinaria (no estocástica, exceptuando el caso en que la condición inicial sea aleatoria)

$$X_t = X_0 + \int_0^t b(s, X_s) ds. \quad (3.5)$$

En la teoría de estas ecuaciones, se acostumbra imponer el supuesto de que $b(t, x)$ satisfaga una condición local de Lipschitz en el espacio de la variable x y que sea acotada sobre subconjuntos compactos de $[0, \infty) \times \mathbb{R}^d$. Estas condiciones aseguran que para $t > 0$ suficientemente pequeño, las iteraciones de Picard-Lindelof [6].

$$X_t^0 \equiv X_0; \quad X_t^{(n+1)} = X_0 + \int_0^t b(s, X_s^{(n)}) ds, \quad n \geq 0, \quad (3.6)$$

converjan a una solución de la ecuación (3.5), y esta solución sea única. En la ausencia de tales condiciones la ecuación podría no tener solución o podría tener un continuo de soluciones. Por ejemplo, la ecuación unidimensional.

$$X_t = \int_0^t |X_s|^\alpha ds, \quad (3.7)$$

tiene una única solución para $\alpha > 1$, a saber, $X_t \equiv 0$, sin embargo, para $0 < \alpha < 1$, todas las funciones de la forma

$$X_t = \begin{cases} 0 & \text{si } 0 \leq t < s, \\ \left(\frac{t-s}{\beta}\right)^\beta & \text{si } s \leq t < \infty. \end{cases} \quad (3.8)$$

con $\beta = \frac{1}{1-\alpha}$ y arbitraria $0 \leq s \leq \infty$, resuelven la ecuación (3.7).

Por esta razón, parece razonable imponer condiciones del tipo Lipchitz para ecuaciones diferenciales estocásticas, e investigar que tipo de resultados de existencia y/o unicidad se pueden obtener bajo estas condiciones. Tales avances fueron desarrollados por primera vez por K.Itô [11].

Supuesto 3.2.1 (Condición global de Lipschitz). Para todo $x, y \in \mathbb{R}^n$ y $t \in [0, T]$, se cumple que

$$|b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| < K|x - y|. \quad (3.9)$$

para alguna constante K , (donde $|\sigma|^2 = \sum |\sigma_{i,j}|^2$).

De la propia definición se deduce que si una función es de Lipschitz en un dominio D , es continua en D . En sentido invertido existen funciones continuas que no son de Lipschitz. Por lo tanto la condición de Lipschitz es una condición más fuerte que la continuidad.

Supuesto 3.2.2 (Crecimiento lineal). Para todo $x, y \in \mathbb{R}^n$ y $t \in [0, T]$ se cumple que

$$|b(t, x)| + |\sigma(t, y)| < C(1 + |x|). \quad (3.10)$$

para alguna constante C , (donde $|\sigma|^2 = \sum |\sigma_{i,j}|^2$).

La condición de crecimiento lineal controla el comportamiento de la solución, es decir el proceso $\mathbf{X} = \{X_t\}_{t \geq 0}$ no explota en un tiempo finito.

A continuación se presentará el resultado principal de este capítulo

Teorema 3.2.3 (Teorema de existencia y unicidad para ecuaciones diferenciales estocásticas). Si las siguientes condiciones se satisfacen:

1. Las funciones $b(t, x)$ y $\sigma(t, x)$ satisfacen los supuestos (3.9), (3.10).
2. $Z = X_0$ es independiente de $B_t = \{B_t\}_{t \geq 0}$ y se cumple que

$$\mathbb{E}(Z^2) < \infty.$$

Entonces la ecuación diferencial estocástica

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \quad 0 \leq t \leq T, \quad X_0 = Z, \quad (3.11)$$

tiene una solución que satisface las siguientes condiciones:

- a) $\mathbf{X} = \{X_t\}_{t \geq 0}$ es continua con probabilidad 1 y $X_0 = Z$ para $t = 0$;
- b) $\sup_{0 \leq t \leq T} \mathbb{E}(X_t^2) < \infty$.

Si X_t^1 y X_t^2 son dos soluciones de que satisfacen a) y b), entonces

$$\mathbb{P}\left(\sup_{0 \leq t \leq T} |X_t^1 - X_t^2| = 0\right) = 1.$$

Los supuestos mencionados anteriormente son naturales en vista de los siguientes dos ejemplos donde $\sigma = 0$.

Ejemplo 3.2.4. La ecuación

$$dX_t = X_t^2 dt, \quad X_0 = 1,$$

correspondiente a $b(x) = x^2$ (la cual no satisface la condición de crecimiento lineal) tiene una solución única

$$X_t = \frac{1}{1-t}; \quad 0 \leq t < 1.$$

Por lo que es imposible encontrar una solución global (definida para toda t) en este caso. La condición de crecimiento lineal asegura que la solución $\mathbf{X} = \{X_t\}_{t \geq 0}$ de la ecuación (3.2.3) no explota, es decir, que $|X_t|$ no tiende a infinito en un tiempo finito.

Ejemplo 3.2.5. La ecuación

$$dX_t = 3X_t^{2/3} dt, \quad X_0 = 0 \tag{3.12}$$

tiene más de una solución. De hecho, para cualquier $a > 0$ la función

$$X_t = \begin{cases} 0 & \text{si } t \leq a \\ (t-a)^3 & \text{si } t > a \end{cases} \tag{3.13}$$

resuelve (3.12). En este caso $b(x) = 3x^{2/3}$ no satisface la condición (3.9) en $x = 0$.

La condición (3.9) garantiza que la ecuación (3.2.3) tiene una única solución, donde unicidad significa que si $\{X_t^1\}_{t \geq 0}$ y $\{X_t^2\}_{t \geq 0}$ son dos procesos continuos tales que

$$\{X_t^1\}_{t \geq 0} = \{X_t^2\}_{t \geq 0} \quad \text{para toda } t \leq T, \text{ casi seguramente.} \tag{3.14}$$

3.3. Soluciones fuertes y débiles

Definición 3.3.1. Una solución fuerte a la ecuación diferencial estocástica (3.2.3) sobre un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ con respecto a movimiento browniano fijo $\mathbf{B} = \{B_t\}_{t \geq 0}$ y condición inicial Z , es un proceso $\mathbf{X} = \{X_t\}_{0 \leq t < \infty}$ con trayectorias muestrales continuas y que cumple las siguientes propiedades:

- 1) $\mathbf{X} = \{X_t\}_{t \geq 0}$ es adaptado a la filtración $\{\mathcal{F}_t\} = \sigma(\mathcal{G} \cup \mathcal{N})$, donde $\mathcal{G} = \sigma(Z, B_s; 0 \leq s \leq t)$; $0 \leq t \leq \infty$ y \mathcal{N} es la colección de conjuntos nulos bajo la medida \mathbb{P} ,
- 2) $\mathbb{P}(X_0 = Z) = 1$,
- 3) $\mathbb{P}\left(\int_0^t \{|b_t(s, X_s)| + \sigma_{i,j}^2(s, X_s)\} ds < \infty\right) = 1$ se cumple para cada $1 \leq i \leq d, 1 \leq j \leq r$ y $0 \leq t < \infty$, y
- 4) La versión integral de (3.2.3)

$$X_t = X_0 + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dB_s; \quad 0 \leq t < \infty, \quad (3.15)$$

se satisface casi seguramente.

El teorema de existencia y unicidad indica que la solución $\mathbf{X} = \{X_t\}_{t \geq 0}$ es de tipo fuerte. Esto esencialmente implica unicidad en las trayectorias. Es posible obtener **soluciones débiles** bajo diferentes supuestos. En muchos casos en estadística, condiciones para soluciones débiles son suficientes puesto que ellas implican que cualesquiera dos soluciones $\mathbf{X}^{(1)} = \{X_t^1\}_{t \geq 0}$ y $\mathbf{X}^{(2)} = \{X_t^2\}_{t \geq 0}$ no son necesariamente idénticas en trayectorias, mientras que sus distribuciones lo son, y esto es suficiente para hacer inferencia. Por supuesto, soluciones fuertes implican ser soluciones débiles pero lo contrario no necesariamente se cumple.

A fin de debilitar de alguna forma los supuestos anteriores asegurando la existencia y solución única de la ecuación (3.2.3), se enunciará un teorema sobre dependencia local sobre los coeficientes de la solución.

Teorema 3.3.2. Supóngase que los coeficientes $a_1(t, x), \sigma_1(t, x), a_2(t, x)$ y $\sigma_2(t, x)$ de las ecuaciones

$$dX_t^i = a_i(t, X_t^i) dt + \sigma_i(t, X_t^i) dB_t, \quad i = 1, 2 \quad (3.16)$$

satisfacen las condiciones del Teorema 3.2.3 y que para algún $N > 0$ con $|x| \leq N$, $a_1(t, x) = a_2(t, x)$ y $\sigma_1(t, x) = \sigma_2(t, x)$. Si $X_t^1 = \{X_t^1\}_{t \geq 0}$ y $X_t^2 = \{X_t^2\}_{t \geq 0}$ son soluciones de la ecuación (3.16) con la misma condición inicial $X_0^1 = X_0^2, \mathbb{E}(X_0^2) < \infty$, y

$\tau_i = \sup\{t \geq 0 \text{ tal que } |X_t^i| \leq N\}$, entonces $\mathbb{P}(\tau_1 = \tau_2) = 1$ y

$$\mathbb{P}\left(\sup_{0 \leq s \leq \tau_1} |X_s^1 - X_s^2| = 0\right) = 1.$$

Demostración. Ver [Gihman & Skorohod 1972].

Observación 3.3.3. Se puede ver que el Teorema 3.3.2 puede ser generalizado al caso donde los coeficientes de las dos ecuaciones coinciden en alguna región arbitraria. Entonces, hasta el momento de su salida de esta región, las soluciones de las dos ecuaciones coincide, siempre que sean iguales en el tiempo.

Teorema 3.3.4. Sean los coeficientes de (3.2.3) medibles para $t \in [0, T]$, $x \in (-\infty, \infty)$ y satisfacen las siguientes condiciones:

1) Para algún K

$$|a(t, x)|^2 + |\sigma(t, x)|^2 \leq K^2(1 + x^2).$$

2) Para cada N existe L_N para el cual

$$|a(t, x) - a(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq L_N|x - y|,$$

con $|x| \leq N, |y| \leq N$. Entonces (3.2.3) tiene una solución $\mathbf{X} = \{X_t\}_{t \geq 0}$ que satisface la condición inicial $X_{t_0} = X_0$, esta solución es única en el sentido del Teorema (3.2.3).

Demostración. Ver [Gihman & Skorohod 1972].

Los teoremas anteriores dan condiciones más generales para existencia y unicidad de las soluciones de ecuaciones diferenciales estocásticas con condiciones iniciales dadas.

Capítulo 4

Procesos de difusión.

En este capítulo se introducirán los procesos centrales que se utilizarán a lo largo de este trabajo, así como algunas de sus propiedades más importantes. Se abordarán desde dos perspectivas diferentes, a saber, como procesos estocásticos solución a una ecuación diferencial estocástica y desde el punto de vista de procesos de Markov que cumplen ciertas características. Este último enfoque será de gran utilidad en capítulos posteriores debido a la probabilidad de transición asociada a un proceso de Markov, la cual es la pieza clave en estimación y simulación.

En muchos problemas de las ciencias naturales a menudo es apropiado considerar sistemas determinísticos donde el estado presente del sistema determina la evolución futura del sistema. Una generalización natural de estos sistemas, son los sistemas estocásticos, para los cuales, el estado actual del sistema determina completamente la probabilidad de ocupar los siguientes estados para todos los tiempos futuros. Tales sistemas son llamados, sistemas de Markov y pueden ser descritos por procesos de Markov.

4.1. Definiciones

Definición 4.1.1. Sea $\mathbf{X} = \{X_t\}_{t \geq 0}$ un proceso estocástico definido en $[0, T]$ y \mathcal{F}_t la mínima σ -álgebra con respecto a la cual cada X_s es medible para $s \leq t$. Suponiendo que existe una función $P(t, x, s, A)$ definida para toda x , $0 \leq t \leq s \leq T$ y conjuntos de Borel A para los cuales se cumple con probabilidad igual a 1 que:

$$\mathbb{P}(X_s \in A | \mathcal{F}_t) = P(t, X_t, s, A) = \mathbb{P}(X_s \in A | X_t), \quad (4.1)$$

para cualesquiera $0 \leq t \leq s \leq T$ y conjunto de Borel A . Entonces el proceso $\mathbf{X} = \{X_t\}_{t \geq 0}$ será llamado un proceso de Markov y la función $P(t, x, s, A)$, la probabilidad de transición para $\mathbf{X} = \{X_t\}_{t \geq 0}$.

De aquí en adelante se supondrá que la probabilidad de transición $P(t, x, s, A)$ para el proceso $\mathbf{X} = \{X_t\}_{t \geq 0}$ satisface las siguientes condiciones.

- (I) $P(t, x, s, A)$ es una función de Borel de x ;
- (II) $P(t, x, s, A)$ es una medida con respecto a A para t, x, s fijas;
- (III) $\int P(t, x, s, dy)P(s, y, \tau, A) = P(t, x, \tau, A)$ para toda x y $0 \leq s < \tau \leq T$.

4.2. Procesos de difusión como procesos de Markov

Definición 4.2.1 (Procesos de difusión). Un proceso de Markov $\mathbf{X} = \{X_t\}_{t \geq 0}$, $t \in \mathbb{R}$ es llamado un proceso de difusión o simplemente una difusión si su función de transición posee las siguientes propiedades:

- 1) para cualquier $\varepsilon > 0$ y $t \in [0, T], x \in (-\infty, \infty)$

$$\lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \int_{|x-y| > \varepsilon} P(t, x, t + \Delta, dy) = 0;$$

- 2) Existen funciones $a(t, x)$ y $b(t, x)$ tales que para toda $\varepsilon > 0, t \in [0, T]$ y $x \in (-\infty, \infty)$

a) $\lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \int_{|x-y| \leq \varepsilon} (y-x)P(t, x, t + \Delta, dy) = a(t, x),$

b) $\lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \int_{|x-y| < \varepsilon} (y-x)^2 P(t, x, t + \Delta, dy) = b(t, x)$

$a(t, x)$ y $b(t, x)$ son llamados los coeficientes de deriva y difusión respectivamente.

La propiedad 1) garantiza que el proceso de difusión no tenga saltos instantáneos.

El inciso a) de la propiedad 2) implica que

$$a(t, x) = \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \mathbb{E}(X_{t+\Delta} - X_t | X_t = x) \quad (4.2)$$

así el coeficiente de deriva $a(t, x)$ es la tasa de cambio instantánea en promedio del procesos dado que $X_t = x$.

Análogamente, el inciso b) de la propiedad 2) implica que el cuadrado del coeficiente de difusión

$$b^2(t, x) = \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \mathbb{E}((X_{t+\Delta} - X_t)^2 | X_t = x), \quad (4.3)$$

denota la tasa instantánea de cambio de las fluctuaciones cuadradas del proceso dado que $X_t = x$.

El nombre “proceso de difusión” se refiere a que bajo ciertas condiciones, es posible describir el movimiento de una partícula difusa, exactamente con su ayuda.

Observación 4.2.2. Para que el proceso $\mathbf{X} = \{X_t\}_{t \geq 0}$ sea una difusión, es suficiente que su probabilidad de transición satisfaga los siguientes supuestos

1*) para algún $\delta > 0$

$$\lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \int |x - y|^{2+\delta} P(t, x, t + \Delta, dy) = 0;$$

2*) Existen funciones $a(t, x)$ y $b(t, x)$ tales que para toda t, x

$$\mathbf{a}^*) \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \int (y - x) P(t, x, t + \Delta, dy) = a(t, x),$$

$$\mathbf{b}^*) \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \int (y - x)^2 P(t, x, t + \Delta, dy) = b(t, x)$$

Las condiciones 1*) y 2*) son en muchas ocasiones más fáciles de verificar.

4.3. Ecuaciones de Kolmogorov

Cuando los coeficientes de deriva y difusión de un proceso de difusión son funciones suaves, entonces su probabilidad de transición $P(t, x, s, y)$ también satisface las siguientes ecuaciones diferenciales parciales.

4.3.1. Ecuación de Kolmogorov hacia adelante

$$\frac{\partial p}{\partial t} + \frac{\partial}{\partial x}(a(t, y)p) - \frac{1}{2} \frac{\partial^2}{\partial y^2}(b^2(t, y)p) = 0, \quad (s, x) \text{ fijos.} \quad (4.4)$$

4.3.2. Ecuación de Kolmogorov hacia atrás

$$\frac{\partial p}{\partial s} + a(s, x) \frac{\partial p}{\partial x} + \frac{1}{2} b^2(s, x) \frac{\partial^2 p}{\partial x^2} = 0, \quad (t, y) \text{ fijos.} \quad (4.5)$$

La ecuación (4.4) se refiere a la evolución hacia adelante con respecto al estado final (t, y) , mientras la ecuación (4.5) se refiere a la evolución hacia atrás con respecto al estado inicial (s, x) . La ecuación (4.4) es comúnmente llamada la ecuación de Fokker-Plank. Ambas ecuaciones se siguen de la ecuación de Chapman-Kolmogorov y de las tres ecuaciones de la Definición 4.2.1.

4.4. Procesos de difusión como soluciones a ecuaciones diferenciales estocásticas

Teorema 4.4.1. Supongamos que $\mathbf{X} = \{X_t\}_{t \geq 0}$ es una solución de

$$dX_t = a(t, X_t)dt + \sigma(t, X_t)dB_t, \quad X_t = x \quad (4.6)$$

cuyos coeficientes satisfacen las condiciones de existencia y unicidad del Capítulo 3. Entonces $\mathbf{X} = \{X_t\}_{t \geq 0}$ es un proceso de Markov cuya probabilidad de transición está definida por

$$P(t, x, s, A) = \mathbb{P}(X_{x,t}(s) \in A), \quad (4.7)$$

donde $X_{x,t}(s)$ es una solución de

$$X_{x,t}(s) = x + \int_t^s a(u, X_{x,t}(u))du + \int_t^s \sigma(u, X_{x,t}(u))dB_u \quad (4.8)$$

sobre el intervalo $[t, T]$.

Demostración. Ver [Gihman & Skorohod 1972].

Teorema 4.4.2. Sea $a(t, x)$ y $\sigma(t, x)$ continua en ambos argumentos y suponiendo que para cada K

$$|a(t, x)|^2 + |\sigma(t, x)|^2 \leq K(1 + |x|^2)$$

y para cada N existe un L_N con $|x| \leq N, |y| \leq N$, para el cual

$$|a(t, x) - a(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq L_n|x - y|.$$

Entonces el proceso $X_t = \{X_t\}_{t \geq 0}$ como solución de la ecuación (4.6) será una difusión con coeficiente de difusión $b(t, x) = \sigma^2(t, x)$ y coeficiente de deriva $a(t, x)$.

Demostración. Ver [Gihman & Skorohod 1972].

Por los dos teoremas anteriores, se demuestra que bajo las condiciones antes mencionadas, un proceso de difusión satisface una ecuación diferencial estocástica.

4.5. El generador infinitesimal de una difusión

Es fundamental para muchas aplicaciones que sea posible asociar un operador A diferencial de segundo orden a un proceso de difusión $\mathbf{X} = \{X_t\}_{t \geq 0}$. La conexión básica entre A y $\{X_t\}_{t \geq 0}$ es que A es el generador del proceso $\{X_t\}_{t \geq 0}$

Definición 4.5.1. Sea $\mathbf{X} = \{X_t\}_{t \geq 0}$ un proceso de difusión en \mathbb{R}^n . El generador infinitesimal A de X_t se define como:

$$Af(x) = \lim_{t \downarrow 0} \frac{\mathbb{E}^x[f(X_t)] - f(x)}{t}; \quad x \in \mathbb{R}^n,$$

siempre que este exista.

El conjunto de funciones $f : \mathbb{R}^n \rightarrow \mathbb{R}$ tal que el límite existe en x se denota como $\mathcal{D}_A(x)$, donde \mathcal{D}_A denota el conjunto de funciones para las cuales el límite existe para todo $x \in \mathbb{R}^n$.

Para encontrar la relación entre el generador infinitesimal A y los coeficientes b, σ en la ecuación diferencial estocástica $dX_t = b(X_t)dt + \sigma(X_t)$ que define a $\{X_t\}_{t \geq 0}$, es necesario conocer la siguiente definición y el siguiente resultado.

Definición 4.5.2 (Tiempo de paro). Sea $\{\mathcal{F}_t\}$ una familia creciente de σ -álgebras. Una función $\tau : \Omega \rightarrow [0, \infty]$ es llamada un tiempo de paro respecto a $\{\mathcal{F}_t\}$ si

$$\{\omega; \tau(\omega) \leq t\} \in \mathcal{F}_t, \quad \text{para todo } t \geq 0.$$

En otras palabras, es posible decidir si el evento $\{\tau \leq t\}$ ha ocurrido, teniendo el conocimiento de la filtración $\{\mathcal{F}_t\}$.

Para encontrar la relación entre el generador infinitesimal A de una difusión $\mathbf{X} = \{X_t\}_{t \geq 0}$ con los coeficientes de deriva y de difusión de una ecuación diferencial estocástica, se necesita el siguiente resultado.

Lema 4.5.3. Sea $\mathbf{Y}^x = \{Y_t\}_{t \geq 0}$ un proceso de Itô sobre \mathbb{R}^n de la forma

$$Y_t^x = x + \int_0^t u(s, \omega) ds + \int_0^t v(s, \omega) dB_s, \quad (4.9)$$

donde $\mathbf{B} = \{B_t\}_{t \geq 0}$ es un movimiento browniano m -dimensional. Sea $f \in C_0^2(\mathbb{R}^n)$ con soporte compacto, y sea τ un tiempo de paro respecto a la filtración $\{\mathcal{F}_t^{(m)}\}$ del movimiento browniano m -dimensional. Asumiendo que $\mathbb{E}^x[\tau] < \infty$ y que tanto $u(t, \omega)$ como

$v(t, \omega)$ son acotadas sobre el conjunto de (t, ω) tal que $Y_t(\omega)$ pertenece al soporte de f . Entonces:

$$\mathbb{E}^x[f(Y_\tau)] = f(x) + \mathbb{E}^x \left[\int_0^\tau \left(\sum_i u_i(s, \omega) \frac{\partial f}{\partial x_i}(Y_s) + \frac{1}{2} \sum_{i,j} (vv^T)_{i,j}(s, \omega) \frac{\partial^2 f}{\partial x_i \partial x_j}(Y_s) \right) ds \right].$$

donde E^x es la esperanza con respecto a la ley de probabilidad natural \mathbb{R}^x para Y_t empezando en x :

$$\mathbb{P}^x[Y_{t_1} \in F_1, \dots, Y_{t_k} \in F_k] = P^0[Y_{t_1}^x \in F_1, \dots, Y_{t_k}^x \in F_k], \quad \text{donde } F_i \text{ son conjuntos de Borel}$$

Demostración. Ver [Oksendal 1997].

El resultado anterior nos da la fórmula para el generador A de un proceso de difusión.

Teorema 4.5.4. Sea $\mathbf{X} = \{X_t\}_{t \geq 0}$ una proceso de difusión que satisface

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t.$$

Si $f \in C_0^2(\mathbb{R}^n)$, entonces $f \in \mathcal{D}_A$ y

$$Af(x) = \sum_i b_i(x) \frac{\partial f}{\partial x_i} + \frac{1}{2} \sum_{i,j} (\sigma \sigma^T)_{i,j}(x) \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

Este resultado será muy útil en la sección de estimación mediante polinomios de Hermite del Capítulo 7.

Demostración. Ver [Oksendal 1997].

Capítulo 5

Algunas familias paramétricas de procesos estocásticos

En este capítulo se presentarán tres de los procesos de difusión más utilizados para el modelaje financiero de activos y tasas de interés con un breve resumen de sus propiedades. Este capítulo está basado en [3] en su totalidad. Además para cada proceso se dará explícitamente su probabilidad de transición, la cual como se mencionó anteriormente es de vital importancia para estimación y simulación, usando el Lema de Itô presentado en el Capítulo 2.

5.1. Proceso de Ornstein-Uhlenbeck

Un proceso $\mathbf{X} = \{X_t\}_{t \geq 0}$ solución a la ecuación diferencial estocástica

$$dX_t = \theta(\mu - X_t)dt + \sigma dB_t, \quad X_0 = x_0, \quad \mu > 0, \quad \sigma > 0. \quad (5.1)$$

se le conoce como proceso de Ornstein-Uhlenbeck. En modelaje financiero σ es interpretado como la volatilidad, μ valor de equilibrio a largo plazo y θ se le conoce como la velocidad de reversión a la media.

Para encontrar la solución de dicha ecuación se aplica el Lema de Itô.

Si elegimos $f(t, x) = e^{\theta t}(x - \mu)$

$$f_t(t, x) = \theta e^{\theta t}(x - \mu), \quad f_x = e^{\theta t}, \quad f_{xx} = 0,$$

entonces,

$$\begin{aligned}
 f(t, X_t) &= e^{\theta t} (X_t - \mu) = f(0, X_0) + \int_0^t f_t(u, x) dt + \int_0^t f_x(u, x) dX_u \\
 &= f(0, X_0) + \int_0^t \theta e^{\theta u} (x - \mu) du \\
 &\quad + \int_0^t e^{\theta u} [\theta(\mu - X_u) du + \sigma dB_u] \\
 &= X_0 - \mu + (X_t - \mu)(e^{\theta t} - 1) + (\mu - X_t)(e^{\theta t} - 1) \\
 &\quad + \sigma \int_0^t e^{\theta u} dB_u,
 \end{aligned}$$

Lo que implica que

$$X_t = (X_0 - \mu)e^{-\theta t} + \mu + \int_0^t e^{-\theta(t-u)} dB_u.$$

Finalmente

$$X_t = X_0 e^{-\theta t} + \mu(1 - e^{-\theta t}) + \int_0^t e^{-\theta(t-u)} dB_u.$$

Este procesos tiene esperanza condicional

$$\begin{aligned}
 \mathbb{E}[X_t | X_0 = x_0] &= \mathbb{E} \left[(x_0 - \mu)e^{-\theta t} + \mu + \int_0^t e^{-\theta(t-u)} dB_u \right] \\
 &= (x_0 - \mu)e^{-\theta t} + \mu + \mathbb{E} \left[\int_0^t e^{-\theta(t-u)} dB_u \right] \\
 &= (x_0 - \mu)e^{-\theta t} + \mu,
 \end{aligned}$$

y varianza condicional

$$\begin{aligned}
 \text{Var}[X_t | X_0 = x_0] &= \mathbb{E} \left[\left(\sigma \int_0^t e^{-\theta(t-u)} dB_u \right)^2 \right] \\
 &= \int_0^t \mathbb{E} \left[\sigma^2 e^{-2\theta(t-u)} du \right] \quad (\text{Isometría de Itô}) \\
 &= \frac{\sigma^2}{2\theta} (1 - e^{-2\theta t}).
 \end{aligned}$$

Por lo tanto

$$X_t | X_0 \sim N \left((x_0 - \mu)e^{-\theta t} + \mu, \frac{\sigma^2}{2\theta} (1 - e^{-2\theta t}) \right).$$

Cuando $t \rightarrow \infty$ se obtiene la distribución estacionaria $X_t | X_0 \sim N \left(0, \frac{\sigma^2}{2\theta} \right)$.

5.2. Modelo de Black-Scholes-Merton o movimiento Browniano geométrico

El modelo de Black-Scholes-Merton es el proceso $\mathbf{X} = \{X_t\}_{t \geq 0}$ solución a la ecuación diferencial estocástica

$$dX_t = \mu X_t dt + \sigma X_t dB_t, \quad X_0 = x_0,$$

con $\sigma > 0$. El parámetro μ es interpretado como la tasa constante de interés y σ como la volatilidad. La solución explícita se puede encontrar haciendo $f(t, x) = \ln(X_t)$, por lo tanto

$$f_t(t, x) = 0, \quad f_x = \frac{1}{x}, \quad f_{xx} = -\frac{1}{x^2}.$$

Entonces,

$$\begin{aligned} f(t, X_t) &= f(0, X_0) + \int_0^t \frac{1}{X_u} dX_u - \frac{1}{2} \int_0^t \frac{1}{X_u^2} (dX_u dX_u) \\ &= f(0, X_0) + \mu t + \sigma B_t - \frac{1}{2} \sigma^2 t \\ &= \ln(x_0) + \left(\mu - \frac{1}{2} \sigma^2\right) t + \sigma B_t \end{aligned}$$

Finalmente

$$X_t = x_0 e^{(\mu - \frac{1}{2} \sigma^2) t + \sigma B_t}.$$

La función de densidad condicional es log-normal con media y varianza del logaritmo de su transformación (es decir, la log-media y log-varianza) están dadas por

$$\mu^* = \log(x_0) + \left(\mu - \frac{1}{2} \sigma^2\right) t, \quad \sigma^{2*} = \sigma^2 t.$$

Recordando que la función generadora de momentos de una variable aleatoria normal $N(\mu, \sigma^2)$ está dada por

$$M_X(s) = \mathbb{E}(e^{sX}) = e^{\mu s + \frac{\sigma^2 s^2}{2}}, \quad -\infty < s < \infty,$$

entonces, se puede calcular los momentos del movimiento Browniano geométrico usando la función generadora de momentos de la distribución normal, es decir,

$$M_{X_t}(s) = \mathbb{E}[e^{sX_t}] = e^{\mu t s + \frac{\sigma^2 t s^2}{2}}, \quad -\infty < s < \infty.$$

Por lo tanto, se puede calcular los primeros dos momentos

$$\begin{aligned}\mathbb{E}[X_t] &= x_0 e^{(\mu + \frac{\sigma^2}{2})t}. \\ \mathbb{E}[X_t^2] &= x_0^2 e^{(2\mu + 2\sigma^2)t}.\end{aligned}$$

Entonces,

$$\text{Var}[X(t)] = x_0^2 e^{2\mu t + \sigma^2 t} (e^{\sigma^2 t} - 1). \quad (5.2)$$

5.3. Modelo de Cox-Ingersoll-Ross

Este modelo fue introducido por Feller como un modelo de crecimiento poblacional y llegó a ser popular en finanzas después de que Cox, Ingersoll y Ross lo propusieran para modelar la tasa corta de interés. El proceso de CIR es la solución a la ecuación diferencial estocástica

$$dX_t = (\theta_1 - \theta_2 X_t)dt + \theta_3 \sqrt{X_t} dB_t, \quad X_0 = x_0 < 0, \quad (5.3)$$

algunas veces parametrizada como

$$dX_t = \theta(\beta - X_t)dt + \sigma \sqrt{X_t} dB_t, \quad X_0 = x_0 > 0, \quad (5.4)$$

donde $\theta_1, \theta_2, \theta_3 \in \mathbb{R}_+$. Si $2\theta_1 > \theta_3^2$, el proceso es estrictamente positivo, de otra forma es no negativo, lo que significa que puede alcanzar el estado 0. Por supuesto, este caso no es admitido en finanzas cuando el proceso de CIR es usado para modelar tasas de interés. La ecuación diferencial estocástica (5.3) tiene una solución explícita

$$X_t = \left(X_0 - \frac{\theta_1}{\theta_2} \right) e^{-\theta_2 t} + \theta_3 e^{-\theta_2 t} \int_0^t e^{\theta_2 u} \sqrt{X_u} dB_u.$$

5.3.0.1. La distribución condicional

Bajo la hipótesis $2\theta_1 > \theta_3^2$, el procesos es también estacionario. Más aún, la probabilidad de transición condicional existe en forma explícita. El proceso $\mathbf{Y} = 2c\mathbf{X}$, con $c = 2\theta_2/\theta_3^2(1 - e^{-\theta_2 t})$, tiene distribución condicional $Y_t|Y_0$, la cual sigue la ley de un distribución χ -cuadrada no centrada con $\nu = 4\theta_1/\theta_3^2$ grados de libertad y parámetro de no centralidad $y_0 e^{-\theta_2 t}$. La densidad de transición del proceso original, (es decir, la distribución de $X_t|X_0 = x_0$ puede ser obtenida de la distribución de $Y_t|Y_0 = y_0$, y el resultado es el siguiente

$$p_\theta(t, y|x_0) = c e^{(-u+\nu)} \left(\frac{u}{\nu}\right)^{q/2} I_q(2\sqrt{uv}),$$

con $u = cx_0e^{-\theta_2 t}$, $c = cy$, $q = 2\theta_1/\theta_3^2 - 1$ donde $I_q(\cdot)$ es la función de Bessel modificada de primer orden de grado q , es decir,

$$I_q(x) = \sum_{k=0}^{\infty} \left(\frac{x}{2}\right)^{2k+q} \frac{1}{k!\Gamma(k+k+1)}, \quad x \in \mathbb{R}$$

y $\Gamma(\cdot)$ es la función Gama (es decir, $\Gamma(z) = \int_0^{\infty} x^{z-1} e^{-x} dx, z \in \mathbb{R}_+$).

Comportamiento de la distribución condicional

Desde el punto de vista de inferencia por verosimilitud, debe notarse que aunque la distribución χ -cuadrada no centrada es una distribución de probabilidad bien conocida, su cálculo no es una tarea sencilla. De hecho, cuando $t \rightarrow \infty$, tanto u como v explotan, lo que significa que cuando la distribución tiene que ser calculada para argumentos realmente grandes, esto implica la evaluación de la distribución χ -cuadrada en la cola derecha. Por otro lado, si $t \rightarrow \infty, u \rightarrow 0$, y $u \rightarrow \frac{2\theta_2 x}{\theta_3^2}$ y por lo tanto se requieren cálculos sobre la cola izquierda de la distribución. Puesto que la distribución χ -cuadrada se basa en la función modificada de Bessel de primer orden $I_q(\cdot)$, se tiene que el problema es esencialmente la evaluación de $I_q(\cdot)$ para argumentos muy pequeños o muy grandes.

El proceso de CIR tiene la misma propiedad de reversión a la media que el proceso de Ornstein-Uhlenbeck. La diferencia entre estos procesos es que la desviación estándar de los cambios en la tasa de interés, en un periodo corto de tiempo, en el modelo de CIR es proporcional a $\sqrt{X_t}$. Por lo tanto para el proceso de CIR, la media de la distribución condicional es la misma que para el proceso de Ornstein-Uhlenbeck

$$\mathbb{E}[X_t | X_0 = x_0] = \frac{\theta_1}{\theta_2} + \left(x_0 - \frac{\theta_1}{\theta_2}\right) e^{-\theta_2 t},$$

y la varianza de la distribución condicional está dada por

$$\text{Var}[X_t | X_0 = x_0] = x_0 \frac{\theta_3^2 (e^{-\theta_2 t} - e^{-2\theta_2 t})}{\theta_2} + \frac{\theta_1 \theta_3^2 (1 - e^{-2\theta_2 t})}{2\theta_2^2}.$$

Capítulo 6

Aproximación de soluciones para ecuaciones diferenciales estocásticas vía simulación

En este capítulo se presentan c métodos de aproximaciones discretas para la simulación de los tres procesos presentados en el Capítulo 5. Se definirá el orden de convergencia de dichas aproximaciones y se hará una comparación entre dichos esquemas de aproximación. Este capítulo está basado totalmente en [4]. Para cada uno de los primeros dos métodos se implementó un código de simulación para los procesos de interés.

Los métodos de simulación normalmente son basados en aproximaciones discretas de la solución continua a una ecuación diferencial estocástica. Los métodos de aproximación son clasificados de acuerdo a sus diferentes propiedades. Principalmente se utilizan dos criterios de optimalidad: Los fuertes y débiles órdenes de convergencia.

6.1. Órdenes de Convergencia

Convergencia de orden fuerte

Una aproximación a tiempo discreto Y_δ de un proceso estocástico continuo Y , con δ el incremento de tiempo máximo de la discretización, se dice que converge fuertemente a Y de orden γ si para cualquier horizonte de tiempo fijo T se satisface que

$$\mathbb{E}[Y_\delta(T) - Y(T)] \leq C\delta^\gamma, \quad \forall \delta < \delta_0,$$

con $\delta_0 > 0$ y C constantes que no dependen de δ .

Convergencia de orden débil

Una aproximación a tiempo discreto Y_δ de un proceso estocástico continuo Y , con δ el incremento de tiempo máximo de la discretización, se dice que converge débilmente a Y de orden β si para cualquier horizonte de tiempo T y para cualquier función g de crecimiento polinomial, continuamente diferenciable de orden $2(\beta + 1)$, se satisface que

$$|\mathbb{E}g(Y(T)) - \mathbb{E}g(Y_\delta(T))| \leq C\delta^\beta, \quad \forall \delta < \delta_0,$$

con δ_0 y C constantes que no dependen de δ .

6.2. Esquema de Euler

Consideremos un proceso de Itô $\mathbf{X} = \{X_t\}_{t_0 \leq t \leq T}$ que satisface la ecuación diferencial estocástica

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)B_t,$$

sobre $t_0 \leq t \leq T$ con condición inicial

$$X_{t_0} = X_0,$$

Para una discretización $t_0 = \tau_0 < \tau_1 < \tau_2 < \dots < \tau_N = T$ del intervalo $[t_0, T]$ **una aproximación de Euler** es un proceso estocástico $Y_t = \{Y_t\}_{t_0 \leq t \leq T}$, con $t \in \mathbb{N}$ que satisface el esquema iterativo

$$Y_{n+1} = Y_n + b(\tau_n, Y_n)(\tau_{n+1} - \tau_n) + \sigma(\tau_n, Y_n)(B_{\tau_{n+1}} - B_{\tau_n}), \quad (6.1)$$

para $n = 0, 1, 2, \dots, N - 1$ con condición inicial

$$Y_0 = X_0,$$

donde se ha escrito

$$Y_n = Y(\tau_n),$$

para el valor de la aproximación al tiempo discretizado τ_n .

Se sabe que la esperanza $\mathbb{E}(\Delta B_n) = 0$ y varianza $\mathbb{E}((\Delta B_n)^2) = \Delta$, donde

$$\Delta B_n = B_{\tau_{n+1}} - B_{\tau_n}. \quad (6.2)$$

Por lo tanto los incrementos (6.2) del proceso de Wiener pueden ser generados por medio de los métodos para generar números pseudo-aleatorios para valores independiente

de la distribución Gaussiana, por el ejemplo el algoritmo Box-Muller o el algoritmo Polar Marsaglia que se pueden ver en [10].

A continuación se hará una implementación del esquema de Euler para simular los tres procesos de difusión presentados en el Capítulo 4. Estos códigos fueron implementados en el lenguaje de programación R.

Ejemplo 6.2.1 (Simulación proceso de Ornstein-Uhlenbeck).

Consideremos el proceso de Ornstein-Uhlenbeck solución a

$$dX_t = (\theta_1 - \theta_2 X_t)dt + \theta_3 dB_t.$$

Para este proceso se tiene que

$$b(t, x) = (\theta_1 - \theta_2 \cdot x) \text{ y } \sigma(t, x) = \theta_3 \cdot x.$$

Por lo tanto, sustituyendo en el esquema iterativo (6.3), el esquema de Euler tiene la siguiente forma:

$$X_{t_{i+1}} = X_{t_i} + (\theta_1 - \theta_2 X_i)\Delta t + \theta_3 X_i \Delta B_t.$$

Código

```
Euler_OU<-function(theta1, theta2, theta3, x0, N) {
  dt<-1/N
  X<-numeric(N+1)
  X[1]=x0
  Z<-rnorm(N)
  for(i in 1:N) {
    X[i+1]=X[i]+(theta1-theta2*X[i])*dt+
            theta3*sqrt(dt)*Z[i]
  }
  Y<-ts(X, start=0, deltat = 1/N)
  return(Y)
}
```

Argumentos

- **N** Número de pasos a simular.
- **theta1** parámetro del proceso .

- **theta2** parámetro del proceso.
- **theta3** parámetro del proceso.
- **x0** Valor inicial.

Ejemplo 6.2.2 (Simulación movimiento browniano geométrico).

Consideremos al movimiento browniano geométrico solución a

$$dX_t = \theta_1 X_t dt + \theta_2 X_t dB_t.$$

Para este proceso se tiene que

$$b(t,x) = (\theta_1 \cdot x) \text{ y } \sigma(t,x) = \theta_2 \cdot x.$$

Por lo tanto, sustituyendo en el esquema iterativo (6.3), el esquema de Euler tiene la siguiente forma:

$$X_{t_{i+1}} = X_{t_i} + \theta_1 X_{t_i} \Delta t + \theta_2 X_{t_i} \Delta B_t.$$

Código

```
Euler_GB<-function(mu, sigma, x0, N) {
  dt<-1/N
  X<-numeric(N+1)
  X[1]=x0
  Z<-rnorm(N)
  for(i in 1:N) {
    X[i+1]=X[i]+mu*X[i]*dt+
    sigma*X[i]*sqrt(dt)*Z[i]
  }
  Y<-ts(X, start=0, deltat = 1/N)
  return(Y)
}
```

Argumentos

- **N** Número de pasos a simular.
- **mu** parámetro del proceso.
- **sigma** parámetro del proceso.

- **x0** Valor inicial.

Ejemplo 6.2.3 (Simulación proceso CIR).

Consideremos el proceso CIR, solución a

$$dX_t = (\theta_1 - \theta_2 X_t)dt + \theta_3 \sqrt{X_t} dB_t, \quad X_0 = x_0.$$

Para este proceso se tiene que

$$b(t, x) = (\theta_1 \cdot x) \text{ y } \sigma(t, x) = \theta_2 \cdot \sqrt{x}.$$

Por lo tanto, sustituyendo en el esquema iterativo (6.3), el esquema de Euler tiene la siguiente forma:

$$X_{t_{i+1}} = X_{t_i} + (\theta_1 - \theta_2 X_i) \Delta t + \theta_3 \sqrt{X_{t_i}} \Delta B_t.$$

Código

```
Euler_CIR<-function(theta1, theta2, theta3, x0, N) {
  dt<-1/N
  X<-numeric(N+1)
  X[1]=x0
  Z<-rnorm(N)
  for(i in 1:N) {
    X[i+1]=X[i]+(theta1-theta2*X[i])*dt+
      sqrt(X[i])*theta3*sqrt(dt)*Z[i]
  }
  Y<-ts(X, start=0, deltat = 1/N)
  return(Y)
}
```

Argumentos

- **N** Número de pasos a simular.
- **theta1** parámetro del proceso.
- **theta2** parámetro del proceso.
- **theta3** parámetro del proceso.
- **x0** Valor inicial

6.3. Esquema de Milstein

El esquema de Milstein hace uso del lema de Itô para incrementar la exactitud de la aproximación añadiendo el término de segundo orden. Si denotamos por σ_x a la derivada parcial de $\sigma(t, x)$ con respecto de x , la aproximación de Milstein tiene la siguiente forma

$$\begin{aligned}
 Y_{n+1} = & Y_n + b(t_n, Y_n)(t_{n+1} - t_n) + \sigma(t_n, Y_n)(B_{n+1} - B_n) \\
 & + \frac{1}{2} \sigma(t_n, Y_n) \sigma_x(t_n, Y_n) ((B_{n+1} - B_n)^2 - (t_{n+1} - t_n)). \tag{6.3}
 \end{aligned}$$

Esta aproximación tiene convergencia débil y fuerte de orden uno.

Ejemplo 6.3.1 (Simulación de proceso de Ornstein-Uhlenbeck por el método de Milstein).

Consideremos el proceso de Ornstein-Uhlenbeck solución a

$$dX_t = (\theta_1 - \theta_2 X_t) dt + \theta_3 dB_t.$$

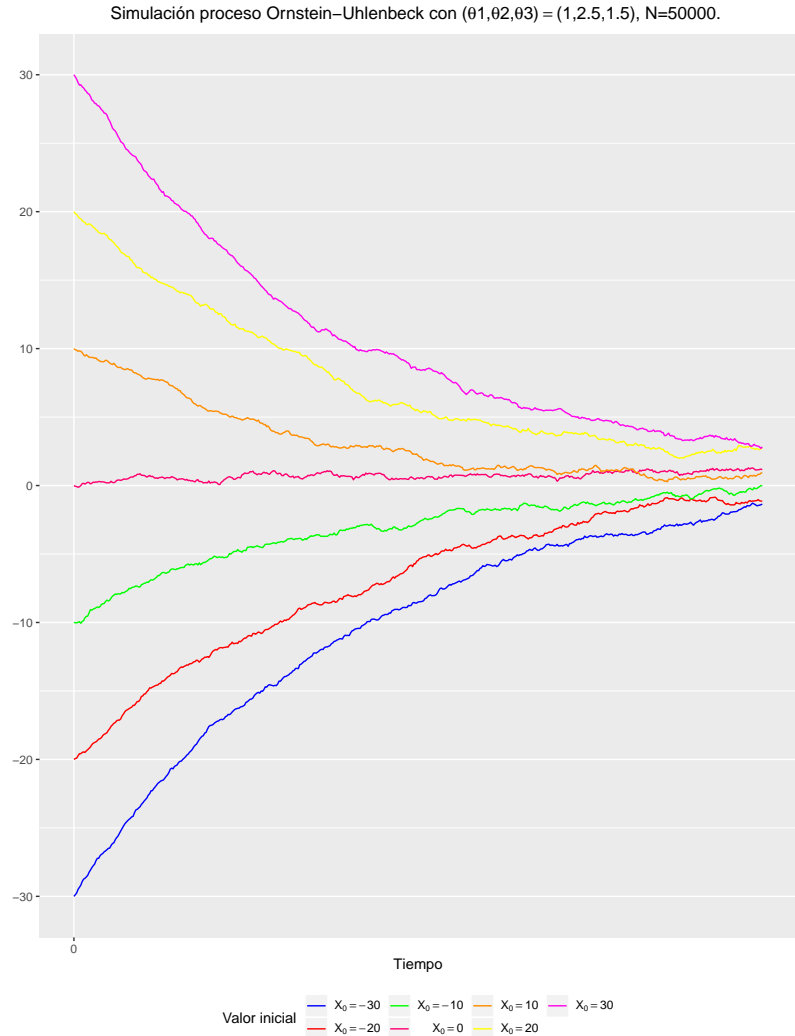
Para este proceso, se tiene que

$$b(t, x) = \theta_1 - \theta_2 \cdot x, \sigma(t, x) = \theta_3.$$

Por lo tanto $\sigma_x(t, x) = 0$.

En este caso el esquema de Milstein y el esquema de Euler coinciden.

A continuación se presentan simulaciones del proceso Orsntein-Uhlenbeck con diferentes valores iniciales. Se puede observar la propiedad de reversión a la media y a diferencia del proceso CIR, este proceso puede tomar valores negativos.



Ejemplo 6.3.2 (Simulación del movimiento browniano geométrico).
 Consideremos el movimiento browniano geométrico solución a

$$dX_t = \theta_1 X_t dt + \theta_2 X_t dB_t.$$

Para este proceso se tiene que

$$b(t, x) = \theta_1 \cdot x, \quad \sigma(t, x) = \theta_2 \cdot x \quad y \quad \sigma_x(t, x) = \theta_2.$$

Por lo tanto, sustituyendo en el esquema iterativo (6.3), el esquema de Milstein tiene la siguiente forma

$$\begin{aligned} X_{t_{i+1}} &= X_{t_i} + \theta_1 X_{t_i} \Delta + \theta_2 X_{t_i} \Delta B_t + \frac{1}{2} \theta_2^2 X_{t_i} ((\Delta B_t)^2 - \Delta t) \\ &= X_{t_i} \left(1 + \left(\theta_1 + \frac{1}{2} \theta_2^2 \right) \Delta t \right) + \theta_2 X_{t_i} \Delta B_t + \frac{1}{2} \theta_2^2 X_{t_i} (\Delta B_t)^2. \end{aligned}$$

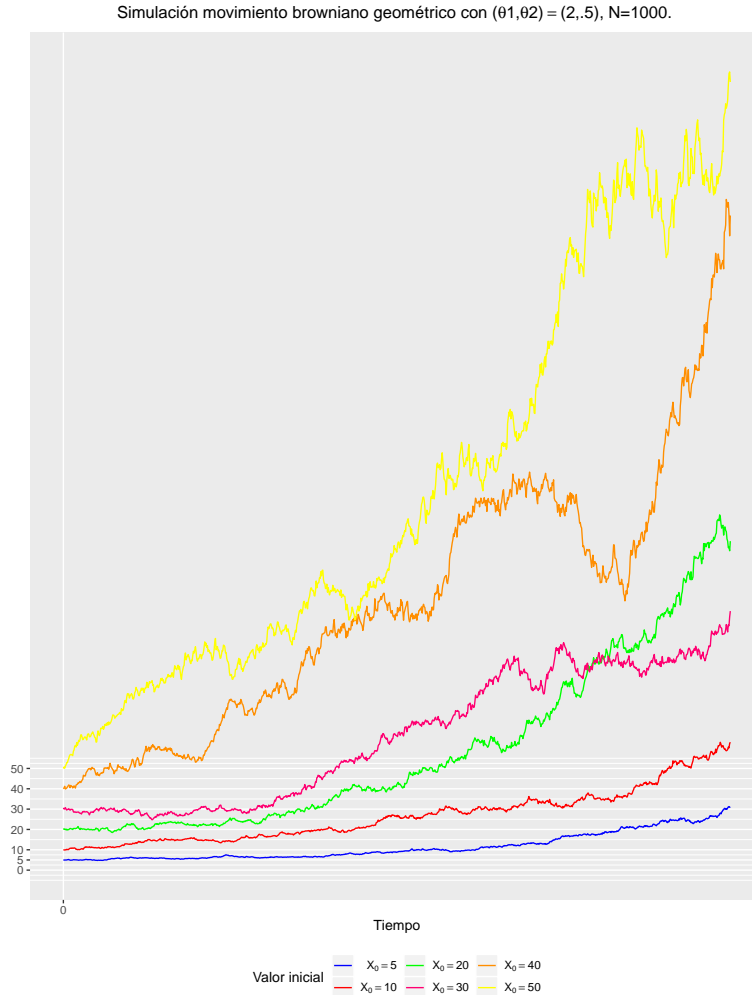
Código

```
Milstein_BG<-function(theta1,theta2,x0,N){
  dt<-1/N
  X<-numeric(N+1)
  X[1]=x0
  Z<-rnorm(N)
  for(i in 1:N){
    X[i+1]=X[i]*(1+(theta[1]-1/2*theta2^2)*dt)+
      theta2*X[i]*sqrt(dt)*Z[i]+1/2*theta2^2*X[i]*
      (sqrt(dt)*Z[i])^2
  }
  Y<-ts(X, start=0,deltat = 1/N)
  return(Y)
}
```

Argumentos

- **N** Número de pasos a simular.
- **mu** parámetro del proceso.
- **sigma** parámetro del proceso.
- **x0** Valor inicial.

A continuación se puede observar simulaciones del movimiento browniano geométrico con diferentes valores iniciales.



Ejemplo 6.3.3 (Simulación del proceso de Cox-Ingersoll-Ross).
 Consideremos el proceso de Cox-Ingersoll-Ross solución a

$$dX_t = (\theta_1 - \theta_2 X_t)dt + \theta_3 \sqrt{X_t} dB_t.$$

Para este proceso se tiene que

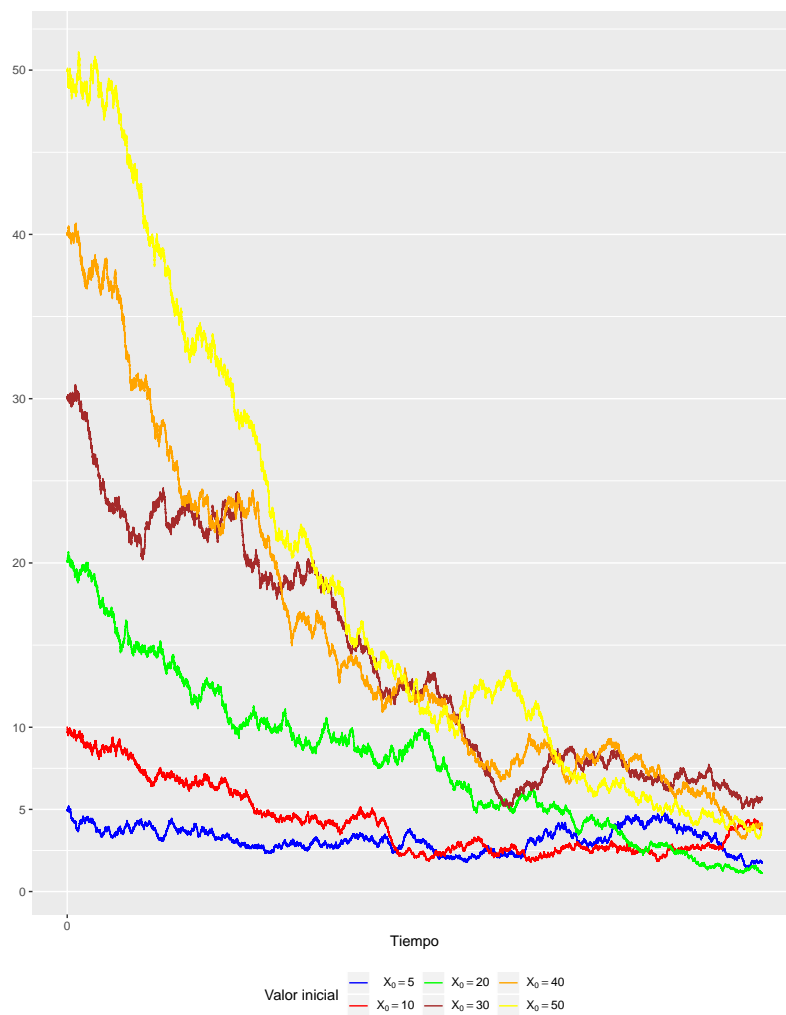
$$b(t, x) = \theta_1 - \theta_2 \cdot x, \sigma(t, x) = \theta_3 \cdot \sqrt{x} \text{ y } \sigma_x(t, x) = \frac{\theta_3}{2\sqrt{x}}.$$

Por lo tanto, sustituyendo en el esquema iterativo (6.3) el esquema de Milstein tiene la siguiente forma

$$X_{t_{i+1}} = X_{t_i} + \theta_1(\theta_2 - X_{t_i})\Delta t + (\theta_3 \sqrt{X_{t_i}})\Delta B_t + \frac{1}{4}\theta_3^2 \left\{ (\Delta B_t)^2 - \Delta t \right\}.$$

A continuación se presentan simulaciones del proceso CIR con diferentes valores iniciales. Recordando que este procesos tiene la propiedad de reversión a la media.

Simulación proceso CIR con $(\theta_1, \theta_2, \theta_3) = (6, 3, 2)$, $N=50000$.



Capítulo 7

Inferencia para procesos de Difusión

El objetivo de este capítulo será hacer estimación paramétrica para los tres procesos de difusión presentados en el Capítulo 5. La primera sección está basada en [3], la segunda sección tiene como fuente principal [7], mientras que la última sección está completamente basada en [5]. Para el primer método se necesitará tener explícitamente la probabilidad de transición del proceso. Para el segundo método se hará el supuesto de que la probabilidad de transición es normal. Para el tercer método se aproximará la probabilidad de transición con una expansión en series.

A lo largo de esta Sección se hará el supuesto de que los procesos son observados a tiempo discreto, puesto que si el proceso se pudiera observar a tiempo continuo de 0 a T, se podrían determinar (en lugar de ser estimados) los parámetros del coeficiente de difusión a través de la variación cuadrática del proceso $\mathbf{X} = \{X_t\}_{t \geq 0}$, y la parte restante podrá ser estimada por máxima verosimilitud: si la función de difusión es completamente conocida, es decir $\sigma(t, X_t) = \sigma(x)$, entonces la función de verosimilitud para $\{X_t\}_{0 \leq t \leq T}$ está dada por

$$L_t^c = \exp \left(\int_0^T \frac{b(t, X_s)}{\sigma^2(X_s)} dX_s - \frac{1}{2} \int_0^T \frac{b^2(t, X_s)}{\sigma^2(X_s)} ds \right).$$

La demostración de este resultado se puede ver en Lipster Shirayev(1977, Capítulo 7).

Consideremos la ecuación diferencial estocástica

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \quad (7.1)$$

donde $\theta \in \Theta \subset \mathbb{R}^p$ es el parámetro multidimensional y θ_0 es el parámetro verdadero para ser estimado. Las funciones $b : \mathbb{R} \times \Omega \rightarrow \mathbb{R}$ y $\sigma : \mathbb{R} \times \Omega \rightarrow (0, \infty)$ son conocidas y tales que la solución de (7.1) existe. El espacio de estados del proceso se denota por $I = (l, r)$, $-\infty \leq l < r \leq +\infty$ es un intervalo abierto y lo mismo para todo θ . Además, para cualquier $\theta \in \Theta$ y cualquier variable aleatoria ξ con soporte en I , la ecuación (7.1) tiene una única solución fuerte para todo $X_0 = \xi$.

Asumimos que el proceso es observado en tiempos discretos $t_i = i\Delta_i, i = 0, 1, \dots, n$, y $T = n\Delta_n$.

Para todo $\theta \in \Theta$ la distribución de \mathbf{X} con $X_0 \sim \xi_\theta$ se denota por p_θ . Además, denotaremos por $p_\theta(t, y|s, x) \equiv P(x, X_{t_0}, X_{t_i}, x_{t_i})$ a la probabilidad de transición de $\mathbf{X} = \{X_t\}_{t \geq 0}$ dado $X_0 = x$. Debido a la propiedad de que los procesos de difusión tienen la propiedad de Markov y por la regla de Bayes, entonces la función de verosimilitud del proceso $\mathbf{X} = \{X_t\}_{t \geq 0}$ condicionado a que $X_{t_0} = x_0$ se puede escribir como el producto de densidades de transición, es decir

$$L_n(\theta) = \prod_{i=1}^n p_\theta(\Delta, X_{t_i} | X_{t_{i-1}}, x_{t_{i-1}}) p_\theta(X_{t_0}). \quad (7.2)$$

7.1. Estimación máxima verosímil

En el método de estimación máxima verosímil asumimos que los datos observados siguen una distribución excepto para un número finito de parámetros desconocidos. Sabemos que la verosimilitud del proceso tiene la siguiente forma

$$L_n(\theta) = \prod_{i=1}^n p_\theta(\Delta, X_{t_i} | X_{t_{i-1}}, x_{t_{i-1}}) p_\theta(X_{t_0}). \quad (7.3)$$

Puesto que la función logaritmo es una función monótona y creciente, es conveniente maximizar la log-verosimilitud que está dada por

$$\ln L(\theta) = \sum_{i=1}^n \ln p_\theta(\Delta, X_{t_i} | X_{t_{i-1}}, x_{t_{i-1}}) p_\theta(X_{t_0}).$$

Para encontrar el máximo de la log-verosimilitud se procede de la siguiente forma

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \sum_{i=1}^n \frac{\partial \ln p_\theta(t, X_{t_{i+1}} | X_{t_i} = x_{t_i})}{\partial \theta} = \Big|_{\theta = \hat{\theta}} = 0. \quad (7.4)$$

Debido a que estamos buscando un máximo es necesario que se cumpla la siguiente desigualdad

$$\frac{\partial^2 \ln L(\theta)}{\partial \theta \partial \theta^T} = \sum_{i=1}^n \frac{\partial^2 \ln p_{\theta}(t, X_{t_{i+1}} | X_{t_i} = x_{t_i})}{\partial \theta \partial \theta^T} \Big|_{\theta = \hat{\theta}} < 0. \quad (7.5)$$

La solución a (7.4), la cual se le conoce como la función score, nos devuelve el parámetro estimado θ .

A continuación se aplicará este método de estimación para los tres procesos de interés en este trabajo.

7.1.1. Ornstein-Uhlenbeck

Suponiendo que el proceso $\{X_t\}_{t \geq 0}$ satisface la ecuación diferencial estocástica

$$dX_t = (\theta_1 - \theta_2 X_t)dt + \theta_3 dB_t, \quad X_0 = x_0. \quad (7.6)$$

Entonces, por el Capítulo 5 sabemos que la probabilidad de transición de este proceso está dada por la siguiente distribución

$$p_{\theta}(t, X_{t_{i+1}} | X_{t_i} = x_{t_i}) \sim N\left(x_{t_i} e^{-\theta_2 t} + \frac{\theta_1}{\theta_2} (1 - e^{-\theta_2 t}), \frac{\theta_3}{2\theta_2} (1 - e^{-2\theta_2 t})\right).$$

Haciendo $\mu = \frac{\theta_1}{\theta_2}$ se tiene que

$$p_{\theta}(t, X_{t_{i+1}} | X_{t_i} = x_{t_i}) \sim N\left(x_{t_i} e^{-\theta_2 t} + \mu (1 - e^{-\theta_2 t}), \frac{\theta_3}{2\theta_2} (1 - e^{-2\theta_2 t})\right).$$

Si $\theta = (\theta_1, \theta_2, \theta_3)$, entonces discretizando las observaciones, la verosimilitud es

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n p_{\theta}(t, X_{t_{i+1}} | X_{t_i} = x_{t_i}) \\ &= \prod_{i=1}^n N\left(x_{t_i} e^{-\theta_2 t} + \mu (1 - e^{-\theta_2 t}), \frac{\theta_3}{2\theta_2} (1 - e^{-2\theta_2 t})\right) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi \frac{\theta_3}{2\theta_2} (1 - e^{-2\theta_2 t})}} \exp\left\{-\frac{(x_{t_{i+1}} - (x_{t_i} e^{-\theta_2 t} + \mu (1 - e^{-\theta_2 t})))^2}{\frac{2\theta_3}{2\theta_2} (1 - e^{-2\theta_2 t})}\right\}, \end{aligned}$$

entonces

$$\begin{aligned} \ln L(\theta) &= \sum_{i=1}^n \ln \left(\frac{1}{\sqrt{2\pi \frac{\theta_3}{2\theta_2} (1 - e^{-2\theta_2 t})}} \exp\left\{-\frac{(x_{t_{i+1}} - (x_{t_i} e^{-\theta_2 t} + \mu (1 - e^{-\theta_2 t})))^2}{\frac{2\theta_3}{2\theta_2} (1 - e^{-2\theta_2 t})}\right\} \right) \\ \ln L(\theta) &= \sum_{i=1}^n -\frac{1}{2} \left[\left(\ln(2\pi) + \ln\left(\frac{\theta_3}{2\theta_2} (1 - e^{-2\theta_2 t})\right) \right) + \frac{(x_{t_{i+1}} - x_{t_i} e^{-\theta_2 t} + \mu (1 - e^{-\theta_2 t}))^2}{\frac{\theta_3}{2\theta_2} (1 - e^{-2\theta_2 t})} \right]. \end{aligned}$$

haciendo $h = \frac{\theta_3}{2\theta_2}(1 - e^{-2\theta_2 t})$, $j = e^{-\theta_2 t}$, $k = \mu(1 - e^{-\theta_2 t})$ obtenemos

$$\ln L(\theta) = -\frac{1}{2} \sum_{i=1}^n \left[\ln(2\pi) + \ln(h) + \frac{(x_{t_{i+1}} - x_{t_i}j + k)^2}{h} \right].$$

Si $\theta = (\mu, \theta_2, \theta_3)$, entonces

Para μ

$$\frac{\partial \ln L(\theta)}{\partial \mu} = \frac{\partial \ln L(\theta)}{\partial k} \frac{\partial k}{\partial \mu} = \sum_{i=1}^n \frac{x_{t_{i+1}} - x_{t_i}j + k}{h} (1 - e^{-\theta_2 t}) = 0.$$

Para θ_3

$$\begin{aligned} \frac{\partial \ln L(\theta)}{\partial \theta_3} &= \frac{\partial \ln L(\theta)}{\partial h} \frac{\partial h}{\partial \theta_3} = \left(-\frac{n}{2h} + \frac{1}{2} \sum_{i=1}^n \frac{(x_{t_{i+1}} - x_{t_i}j + k)^2}{h^2} \right) \\ &\quad \times \left(\frac{1 - e^{-2\theta_2 t}}{2\theta_2} \right) = 0. \end{aligned}$$

Para θ_2

$$\begin{aligned} \frac{\partial \ln L(\theta)}{\partial \theta_2} &= \frac{\partial \ln L(\theta)}{\partial h} \frac{\partial h}{\partial \theta_2} + \frac{\partial \ln L(\theta)}{\partial j} \frac{\partial j}{\partial \theta_2} + \frac{\partial \ln L(\theta)}{\partial k} \frac{\partial k}{\partial \theta_2} \\ &= \left(-\frac{n}{2h} + \frac{1}{2} \sum_{i=1}^n \frac{(x_{t_{i+1}} - x_{t_i}j + k)^2}{h^2} \right) \left(\frac{-\theta_3 e^{-2\theta_2 t} (-2\theta_2 t + e^{2\theta_2 t} - 1)}{2\theta_2^2} \right) \\ &\quad + \left(\sum_{i=1}^n \frac{x_{t_{i+1}}x_{t_i} - jx_{t_i}^2 - kx_{t_i}}{h} \right) (-te^{-\theta_2 t}) + \sum_{i=1}^n \frac{x_{t_{i+1}} - x_{t_i}j - k}{h} (\mu e^{-\theta_2 t} t) = 0. \end{aligned}$$

Entonces necesariamente

$$\sum_{i=1}^n \frac{x_{t_{i+1}} - x_{t_i}j - k}{h} = 0 \quad (7.7)$$

$$-\frac{n}{2h} + \frac{1}{2} \sum_{i=1}^n \frac{(x_{t_{i+1}} - x_{t_i}j + k)^2}{h^2} = 0 \quad (7.8)$$

$$\sum_{i=1}^n \frac{x_{t_{i+1}}x_{t_i} - jx_{t_i}^2 - kx_{t_i}}{h} = 0 \quad (7.9)$$

Despejando h de la expresión (7.8), se tiene que

$$\begin{aligned}
 & -\frac{n}{2h} + \frac{1}{2} \sum_{i=1}^n \frac{(x_{t_{i+1}} - x_{t_i}j + k)^2}{h^2} = 0 \\
 \Leftrightarrow & \frac{n}{2h} = \frac{1}{2} \sum_{i=1}^n \frac{(x_{t_{i+1}} - x_{t_i}j + k)^2}{h^2} \\
 \Leftrightarrow & \frac{2h^2n}{2h} \sum_{i=1}^n (x_{t_{i+1}} - x_{t_i}j + k)^2 \\
 \Leftrightarrow & h = \frac{\sum_{i=1}^n (x_{t_{i+1}} - x_{t_i}j + k)^2}{n}. \tag{7.10}
 \end{aligned}$$

Sustituyendo la expresión (7.10) en la ecuación (7.7) y puesto que $h \neq 0$ se tiene que

$$\begin{aligned}
 & \sum_{i=1}^n \frac{x_{t_{i+1}} - jx_{t_i} - k}{h} = 0 \\
 \Leftrightarrow & \sum_{i=1}^n x_{t_{i+1}} - \sum_{i=1}^n jx_{t_i} - nk = 0 \\
 \Leftrightarrow & nk = \sum_{i=1}^n x_{t_{i+1}} - \sum_{i=1}^n jx_{t_i} \\
 \Leftrightarrow & k = \frac{\sum_{i=1}^n x_{t_{i+1}} - \sum_{i=1}^n jx_{t_i}}{n}. \tag{7.11}
 \end{aligned}$$

Por último, sustituyendo la expresión (7.11) en la ecuación (7.9) y puesto que $h \neq 0$, se tiene que

$$\begin{aligned}
 & \sum_{i=1}^n x_{t_{i+1}}x_{t_i} - jx_{t_i}^2 - kx_{t_i} = 0 \\
 \Leftrightarrow & \sum_{i=1}^n x_{t_{i+1}}x_{t_i} - \sum_{i=1}^n jx_{t_i}^2 - \sum_{i=1}^n kx_{t_i} = 0 \\
 \Leftrightarrow & \sum_{i=1}^n x_{t_{i+1}}x_{t_i} - \sum_{i=1}^n jx_{t_i}^2 - \left(\sum_{i=1}^n x_{t_i}\right) \left(\frac{\sum_{i=1}^n x_{t_{i+1}} - \sum_{i=1}^n jx_{t_i}}{n}\right) = 0 \\
 \Leftrightarrow & \sum_{i=1}^n x_{t_{i+1}}x_{t_i} - \sum_{i=1}^n jx_{t_i}^2 - \sum_{i=1}^n x_{t_i} \frac{\sum_{i=1}^n x_{t_{i+1}}}{n} + \sum_{i=1}^n x_{t_i} \frac{\sum_{i=1}^n jx_{t_i}}{n} = 0 \\
 \Leftrightarrow & j \left(\sum_{i=1}^n x_{t_i}^2 - \frac{\left(\sum_{i=1}^n x_{t_i}\right)^2}{n} \right) = \sum_{i=1}^n x_{t_{i+1}}x_{t_i} - \frac{\sum_{i=1}^n x_{t_{i+1}}x_{t_i}}{n} \\
 \Leftrightarrow & j = \frac{\sum_{i=1}^n x_{t_{i+1}}x_{t_i} - \frac{\sum_{i=1}^n x_{t_{i+1}}x_{t_i}}{n}}{\left(\sum_{i=1}^n x_{t_i}\right)^2} \\
 & \quad \frac{\sum_{i=1}^n x_{t_i}^2 - \frac{\left(\sum_{i=1}^n x_{t_i}\right)^2}{n}}{n} \\
 \Leftrightarrow & j = \frac{n \sum_{i=1}^n x_{t_{i+1}}x_{t_i} - \sum_{i=1}^n x_{t_{i+1}}x_{t_i}}{n \sum_{i=1}^n x_{t_i}^2 - \left(\sum_{i=1}^n x_{t_i}\right)^2}. \tag{7.12}
 \end{aligned}$$

Por lo tanto, se tienen las expresiones para h , j y k .

$$\begin{aligned}
 j &= \frac{n \sum_{i=1}^n x_{t_{i+1}}x_{t_i} - \sum_{i=1}^n x_{t_{i+1}} \sum_{i=1}^n x_{t_i}}{n \sum_{i=1}^n x_{t_i}^2 - \left(\sum_{i=1}^n x_{t_i}\right)^2} \\
 k &= \frac{\sum_{i=1}^n x_{t_i} - j \sum_{i=1}^n x_{t_{i+1}}}{n} \\
 h &= \frac{\sum_{i=1}^n (x_{t_{i+1}} - jx_{t_i} - k)^2}{n}. \tag{7.13}
 \end{aligned}$$

De los cambios que se hicieron para j , k , h .

$$h = \frac{\theta_3}{2\theta_2}(1 - e^{-2\theta_2 t}), \quad j = e^{-\theta_2 t}, \quad k = \mu(1 - e^{-\theta_2 t}), \quad \mu = \frac{\theta_1}{\theta_2}$$

se pueden obtener los estimadores máximos verosímiles, que son los siguientes

$$\begin{aligned}\hat{\theta}_2 &= -\frac{\ln(j)}{t}, \\ \hat{\theta}_3 &= \frac{h(2\hat{\theta}_2)}{1 - e^{-2\hat{\theta}_2}}, \\ \hat{\theta}_1 &= \sqrt{\frac{k}{1-j}}\hat{\theta}_2.\end{aligned}$$

A continuación se muestra como funciona este algoritmo para diferente número de trayectorias del proceso de Ornstein-Uhlenbeck, donde se calcula la media muestra, la varianza muestral e intervalo de confianza al 95 % con la función `Vasicek_media_varianza_intervalos()` descrita en el apéndice A.1.3.

Ejemplo

```
Vasicek_media_varianza_intervalos(1000, 1, 2, 0.5, .5, 1000, 500, 200, 100)
```

```
## $Theta1
##           Media      Varianza      25%      97.5%
## 1000 1.000224 0.007292155 0.9409889 1.179455
## 500  1.000401 0.006650217 0.9438583 1.167390
## 200  1.003935 0.006040820 0.9448134 1.163413
## 100  1.001366 0.006703656 0.9426991 1.167390
##
## $Theta2
##           Media      Varianza      25%      97.5%
## 1000 2.004834 0.02732061 1.889324 2.361014
## 500  2.003604 0.02509277 1.899179 2.310590
## 200  2.007870 0.02161366 1.900713 2.308102
## 100  2.005746 0.02414132 1.900582 2.333086
##
## $Theta3
##           Media      Varianza      25%      97.5%
## 1000 0.4994921 0.0002957658 0.4877125 0.5316149
## 500  0.4999260 0.0002852722 0.4879715 0.5291107
## 200  0.5016762 0.0002378379 0.4903227 0.5276027
## 100  0.5019316 0.0002648959 0.4898587 0.5276572
```

7.1.2. Modelo de Black-Scholes-Merton

Supongamos que el proceso estocástico satisface la ecuación diferencial estocástica

$$dX_t = \mu X_t dt + \sigma X_t dB_t.$$

Por el Capítulo 5 se sabe que la probabilidad de transición está dada por

$$p_{\theta}(t, y|x_0) = \frac{1}{y\sigma\sqrt{2\pi t}} \exp\left\{\frac{-(\log y - \mu)^2}{2\sigma^2 t}\right\}. \quad (7.14)$$

Discretizando las observaciones se tiene que la verosimilitud es

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n p_{\theta}(t, X_{t_{i+1}}|X_{t_i} = x_{t_i}) \\ &= \prod_{i=1}^n \left[\frac{1}{x_{t_i}\sqrt{2\pi\sigma}} \exp\left\{\frac{-(\ln(x_{t_i}) - \mu)^2}{2\sigma^2 t}\right\} \right] \\ &= (2\pi\sigma^2 t)^{-n/2} \prod_{i=1}^n \frac{1}{x_{t_{i+1}}} \exp\left\{\sum_{i=1}^n \frac{-\ln(x_{t_i}) - \mu)^2}{2\sigma^2 t}\right\}. \\ \Rightarrow \ln L(\theta) &= \frac{-n}{2} \ln(2\pi\sigma^2 t) - \sum_{i=1}^n \ln(x_{t_i}) - \frac{\sum_{i=1}^n (\ln(x_{t_i}) - \mu)^2}{2\sigma^2 t} \\ &= \frac{-n}{2} \ln(2\pi\sigma^2 t) - \sum_{i=1}^n \ln(x_{t_i}) - \frac{\sum_{i=1}^n (\ln(x_{t_i})^2 - 2\ln(x_{t_i})\mu + \mu^2)}{2\sigma^2 t} \\ &= \frac{-n}{2} \ln(2\pi\sigma^2 t) - \sum_{i=1}^n \ln(x_{t_i}) - \frac{\sum_{i=1}^n \ln(x_{t_i})^2}{2\sigma^2 t} + \mu \frac{\sum_{i=1}^n \ln(x_{t_i})}{\sigma^2 t} - \frac{\sum_{i=1}^n \mu^2}{2\sigma^2 t} \\ &= \frac{-n}{2} \ln(2\pi\sigma^2 t) - \sum_{i=1}^n \ln(x_{t_i}) - \frac{\sum_{i=1}^n \ln(x_{t_i})^2}{2\sigma^2 t} + \mu \frac{\sum_{i=1}^n \ln(x_{t_i})}{\sigma^2 t} - \frac{n\mu^2}{2\sigma^2 t}. \end{aligned} \quad (7.15)$$

Para encontrar el máximo se toma el gradiente de (7.15) y se iguala a cero

Para μ

$$\begin{aligned} \frac{\partial \ln L(\theta)}{\partial \mu} &= \frac{\sum_{i=1}^n \ln(x_{t_i})}{\sigma^2 t} - \frac{2n\mu}{2\sigma^2 t} = 0 \\ \Rightarrow \frac{n\mu}{\sigma^2 t} &= \frac{\sum_{i=1}^n \ln(x_{t_i})}{\sigma^2 t} \\ \Rightarrow \hat{\mu} &= \frac{\sum_{i=1}^n \ln(x_{t_i})}{n}. \end{aligned}$$

Para σ^2

$$\frac{\partial \ln L(\theta)}{\sigma^2} = -\frac{n}{2\sigma^2 t} - \frac{\sum_{i=1}^n (\ln(x_{t_i}) - \mu)^2}{2t} \cdot \frac{1}{(-\sigma^2)^2} = 0 \quad (7.16)$$

$$\Rightarrow \frac{n}{2\sigma^2 t} = \frac{\sum_{i=1}^n (\ln(x_{t_i}) - \mu)^2}{2\sigma^4 t} \quad (7.17)$$

$$\Rightarrow n = \frac{\sum_{i=1}^n (\ln(x_{t_i}) - \mu)^2}{\sigma^2 t} \quad (7.18)$$

$$\Rightarrow \sigma^2 = \frac{\sum_{i=1}^n (\ln(x_{t_i}) - \hat{\mu})^2}{nt}. \quad (7.19)$$

7.1.3. Proceso de Cox-Ingersoll-Ross

Recordando que el proceso de Cox-Ingersoll-Ross es el proceso solución a la ecuación diferencial estocástica

$$dX_t = (\kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dB_t), \quad X_0 = x_0 > 0, \quad (7.20)$$

La densidad de transición para este modelo fue derivada por Feller y está dada por

$$L(\theta) = \prod_{i=1}^n c e^{(-u_i + v_{i+1})} \left(\frac{v_{i+1}}{u_i} \right)^{\frac{q}{2}} I_q(2\sqrt{u_i v_{i+1}}) \quad (7.21)$$

donde

$$\begin{aligned} c &= \frac{2\kappa}{\sigma^2(1 - e^{-\kappa\Delta})} \\ u_i &= cX_i e^{-\kappa\Delta t}, \\ v_{i+1} &= cX_{i+1}, \\ q &= \frac{2\kappa\alpha}{\sigma^2} - 1. \end{aligned}$$

y $I_q(\cdot)$ es la función modificada de Bessel de primer orden y orden q

Entonces la log-verosimilitud está dada por

$$\ln L(\theta) = n \ln c + \sum_{i=1}^n \left\{ -u_i - v_{i+1} + \frac{q}{2} \ln \left(\frac{v_{i+1}}{u_i} \right) + \ln \left(I_q(2\sqrt{u_i v_{i+1}}) \right) \right\}. \quad (7.22)$$

7.2. Estimación máxima verosímil por el método de Euler-Maruyama

Como se vio en el Capítulo 6, la discretización de Euler-Maruyama para una ecuación diferencial estocástica

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t, \quad (7.23)$$

está dada por

$$X_{t_{i+1}} = X_{t_i} + \mu(t, X_{t_i})(t_{i+1} - t_i) + \sigma(t, X_{t_i})\sqrt{t_{i+1} - t_i} \cdot N(0, 1). \quad (7.24)$$

Este método de discretización alcanza un orden de convergencia de 0.5, si se cumplen las condiciones de Lipschitz y de crecimiento acotado, a la solución de la ecuación diferencial estocástica (7.23) cuando $\Delta \rightarrow 0$. Aún más, cuando $\sigma(X_t; \theta) = \sigma(\theta)$, este método puede llegar a alcanzar órdenes de convergencia más altos. Por medio de la relación iterativa (7.24) se puede simular el proceso $\mathbf{X} = \{X_t\}_{t \geq 0}$ para un valor inicial $X_0 = x_0$ dado.

En este método de estimación se supondrá que la densidad de transición no es conocida.

A continuación se mostrará como funciona este método.

Dadas $n + 1$ observaciones de un proceso de difusión, considérese los datos X_t muestreado en fechas no estocásticas $t_0 = 0 < t_1 < t_2 < \dots < t_n$.

La densidad conjunta de la muestra está dada por

$$L_n(\theta) = \prod_{i=1}^n p_\theta(\Delta, X_{t_i} | X_{t_{i-1}}) p_\theta(X_0).$$

donde $p_\theta(X_0)$ es la densidad de marginal de X_0 y $p_\theta(\Delta, X_{t_i} | X_{t_{i-1}})$ representan las densidades de transición. En esta sección se examinará como se puede abordar este problema utilizando la distribución Gaussiana para aproximar la probabilidad de transición.

Cuando la distancia entre las observaciones es suficientemente pequeña, tales aproximaciones conducen a razonables (aunque sesgados) estimadores. Una discusión sobre esto puede verse en [9].

Considérese la ecuación diferencial estocástica

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t, \quad (7.25)$$

Asumiendo que las funciones $\mu(t, X_t)$ y $\sigma(t, X_t)$ son conocidas, excepto para el parámetro θ , el cuál varía en un subconjunto $\Theta \in \mathbb{R}^p$. Se discretiza el proceso solución a la

ecuación (7.25), suponiendo que los coeficientes de deriva y de difusión son constantes en el intervalo de tiempo $\Delta_i = t_{i+1} - t_i$, como sigue:

$$X_{t_{i+1}} - X_{t_i} = \mu(t, X_{t_i})(t_{i+1} - t_i) + \sigma(t, X_{t_i})(B_{t_{i+1}} - B_{t_i}). \quad (7.26)$$

Puesto que $B_{t_{i+1}} - B_{t_i} | B_{t_i} \sim N(0, \sigma^2(t, X_{t_i})\Delta_i)$, de hecho se está asumiendo que:

$$\begin{aligned} \mathbb{E}(X_{t_{i+1}} | X_{t_i}) &= \mu(t, X_{t_i})\Delta_i + X_{t_i} \\ \mathbb{E}\left[(X_{t_{i+1}} - \mathbb{E}(X_{t_{i+1}} | X_{t_i}))^2 | X_{t_i}\right] &= \sigma^2(t, X_{t_i})\Delta_i. \end{aligned}$$

Por lo tanto la distribución condicional del proceso es la siguiente

$$X_{t_{i+1}} | X_{t_i} \sim N(X_{t_i} + \mu(t, X_{t_i})\Delta, \sigma^2(t, X_{t_i})\Delta). \quad (7.27)$$

y por lo tanto, la función de verosimilitud está dada por:

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n p_{\theta}(t, X_{t_{i+1}} | X_{t_i} = x_{t_i}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2(t, X_{t_i})\Delta}} \\ &\quad \times \exp\left\{-\frac{(X_{t_{i+1}} - (X_{t_i} + \mu(t, X_{t_i})\Delta))^2}{2\sigma^2(t, X_{t_i})\Delta}\right\}. \end{aligned} \quad (7.28)$$

esto implica que la función de log-verosimilitud para este método de discretización es de la forma

$$\ln L(\theta) = -\frac{n}{2}(2\pi) - \frac{1}{2} \sum_{i=1}^n \ln(\sigma^2(t, X_{t_i})\Delta) - \frac{1}{2} \sum_{i=1}^n \frac{(X_{t_{i+1}} - (X_{t_i} + \mu(t, X_{t_i})\Delta))^2}{\sigma^2(t, X_{t_i})\Delta}. \quad (7.29)$$

Al maximizar la función de log-verosimilitud con respecto a θ se produce la siguiente función (score).

$$\begin{aligned} U(\theta; X) &= \sum_{i=1}^n \left[\left(X_{t_{i+1}} - (X_{t_i} + \mu(t, X_{t_i})\Delta) \right)^2 \frac{\partial_{\theta} \sigma^2(t, X_{t_i})}{2\sigma^4(t, X_{t_i})\Delta} \right. \\ &\quad \left. - \frac{1}{2} \left(\frac{\partial_{\theta} \sigma^2(t, X_{t_i})}{\sigma^2(t, X_{t_i})} \right) + \frac{(X_{t_{i+1}} - [X_{t_i} + \mu(t, X_{t_i})\Delta])}{\sigma^2(t, X_{t_i})} \partial_{\theta} \mu(t, X_{t_i}) \right] \end{aligned} \quad (7.30)$$

$$\begin{aligned} &= \sum_{i=1}^n \left[\frac{\partial_{\theta} \sigma^2(t, X_{t_i})}{2\sigma^4(t, X_{t_i})\Delta} \left((X_{t_{i+1}} - [X_{t_i} + \mu(t, X_{t_i};) \Delta])^2 - \sigma^2(t, X_{t_i}) \Delta \right) \right. \\ &\quad \left. + \frac{\partial_{\theta} \mu(t, X_{t_i})}{\sigma^2(t, X_{t_i})} (X_{t_{i+1}} - [X_{t_i} + \mu(t, X_{t_i})\Delta]) \right]. \end{aligned} \quad (7.31)$$

Por lo tanto, el parámetro desconocido θ puede ser calculado explícitamente o numéricamente igualando la expresión (7.31) a cero.

7.2.1. Ornstein-Uhlenbeck

Considérese la ecuación diferencia estocástica

$$dX_t = \theta(\mu - X_t)dt + \sigma dB_t, \quad X_0 = x_0. \quad (7.32)$$

Aplicando el método de discretización de Euler-Maruyama para el modelo de Ornstein-Uhlenbeck se tiene que

$$X_{t_{i+1}} = X_{t_i} + \theta(\mu - X_{t_i})\Delta + \sigma\sqrt{\Delta}Z; \quad \text{donde } Z \sim N(0, 1). \quad (7.33)$$

Entonces la función de verosimilitud se puede obtener sustituyendo $\mu(t, X_{t_i}) = \theta(\mu - X_{t_i})$ y $\sigma^2(t, X_{t_i}) = \sigma^2$ en (7.28), es decir,

$$L(\theta) = \prod_{i=1}^n \frac{1}{2\pi\sigma^2\Delta} \exp \left\{ -\frac{(X_{t_{i+1}} - (X_{t_i} + \theta(\mu - X_{t_i})\Delta))^2}{2\sigma^2\Delta} \right\}. \quad (7.34)$$

Por lo tanto, la log-verosimilitud está dada por:

$$\ln L(\theta) = -\frac{n}{2} \ln(2\pi\sigma^2\Delta) - \sum_{i=1}^n \frac{(X_{t_{i+1}} - X_{t_i}(1 - \theta\Delta) - \theta\mu\Delta)^2}{2\sigma^2\Delta}. \quad (7.35)$$

Para θ

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \frac{1}{\sigma^2} \sum_{i=1}^n (\mu - X_{t_i})(X_{t_{i+1}} - X_{t_i} - \theta(\mu - X_{t_i})\Delta).$$

Para μ

$$\frac{\partial \ln L(\theta)}{\partial \mu} = \frac{\theta}{\sigma^2} \sum_{i=1}^n (X_{t_{i+1}} - X_{t_i} - \theta(\mu - X_{t_i})\Delta).$$

Para σ

$$\frac{\partial \ln L(\theta)}{\partial \sigma} = \frac{1}{\sigma^3} \sum_{i=1}^n (X_{t_{i+1}} - X_{t_i} - \theta(\mu - X_{t_i})\Delta)^2 - \frac{n}{\sigma}.$$

igualando a cero cada ecuación se tiene que necesariamente

$$\sum_{i=1}^n (\mu - X_{t_i})(X_{t_{i+1}} - X_{t_i} - \theta(\mu - X_{t_i})\Delta) = 0. \quad (7.36)$$

$$\sum_{i=1}^n (X_{t_{i+1}} - X_{t_i} - \theta(\mu - X_{t_i})\Delta) = 0. \quad (7.37)$$

$$\sum_{i=1}^n (X_{t_{i+1}} - X_{t_i} - \theta(\mu - X_{t_i})\Delta)^2 - n\sigma^2\Delta = 0. \quad (7.38)$$

Despejando la ecuación (7.38) se tiene que

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n (X_{t_{i+1}} - X_{t_i} - \hat{\theta}(\hat{\mu} - X_{t_i})\Delta)^2}{n\Delta}}. \quad (7.39)$$

De la ecuación (7.37) se tiene que

$$\begin{aligned} \sum_{i=1}^n (X_{t_{i+1}} - X_{t_i} - \theta(\mu - X_{t_i})\Delta) &= 0 \\ \iff \sum_{i=1}^n X_{t_{i+1}} - \sum_{i=1}^n X_{t_i} - \sum_{i=1}^n \mu\theta\Delta + \sum_{i=1}^n \theta\Delta X_{t_i} &= 0 \\ \iff \hat{\mu} &= \frac{\sum_{i=1}^n X_{t_{i+1}} - \sum_{i=1}^n X_{t_i} + \theta\Delta \sum_{i=1}^n X_{t_i}}{n\theta\Delta}. \end{aligned} \quad (7.40)$$

Sustituyendo μ en la ecuación (7.36) se tiene que

$$\hat{\theta} = \frac{\sum_{i=1}^n X_{t_i}^2 + \frac{1}{n} \sum_{i=1}^n X_{t_i} (X_{t_n} - X_{t_1}) - \sum_{i=1}^n X_{t_i} X_{t_{i+1}}}{\Delta \left(\sum_{i=1}^n X_{t_i}^2 - \frac{1}{n} \left(\sum_{i=1}^n X_{t_i} \right)^2 \right)}. \quad (7.41)$$

7.2.2. Proceso de Cox-Ingersoll-Ross

Considérese la ecuación diferencia estocástica

$$dX_t = (\kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dB_t), \quad X_0 = x_0 > 0, \quad (7.42)$$

Usando el método de discretización de Euler-Maruyama para esta ecuación diferencial estocástica se obtiene que

$$X_{t_{i+1}} = X_{t_i} + \kappa(\alpha - X_{t_i})\Delta + \sigma\sqrt{X_{t_i}}\sqrt{\Delta}Z, \quad \text{donde } Z \sim N(0, 1). \quad (7.43)$$

Para este modelo se tiene que $\mu(t, X_{t_i}) = \kappa(\alpha - X_{t_i})$ y $\sigma^2(t, X_{t_i}) = \sigma^2 X_{t_i}$.

Se tiene que la función de verosimilitud se puede obtener sustituyendo $\mu(t, X_{t_i}) = \theta(\mu - X_{t_i})$ y $\sigma^2(t, X_{t_i}) = \sigma^2$ en la expresión (7.28), es decir, la verosimilitud obtiene la siguiente forma

$$L(\theta) = \prod_{i=1}^n \frac{1}{2\pi\sigma^2 X_{t_i}\Delta} \exp\left\{-\frac{X_{t_{i+1}} - (X_{t_i} + \kappa(\alpha - X_{t_i})\Delta))^2}{2\sigma^2 X_{t_i}\Delta}\right\}. \quad (7.44)$$

Por lo tanto, la log-verosimilitud está dada por:

$$\ln L(\theta) = -\frac{n}{2} \ln(2\pi\sigma^2 X_{t_i}\Delta) - \sum_{i=1}^n \frac{(X_{t_{i+1}} - X_{t_i}(1 - \kappa\Delta - \kappa\alpha\Delta))^2}{2\sigma^2 X_{t_i}\Delta}. \quad (7.45)$$

En consecuencia, se puede obtener la matriz (score) de la expresión (7.31) para estimar los parámetros.

Por lo tanto

Para κ

$$\frac{\partial \ln L(\theta)}{\partial \kappa} = \frac{1}{\sigma^2} \sum_{i=1}^n \frac{\alpha - X_{t_i}}{X_{t_i}} (X_{t_{i+1}} - X_{t_i} - \kappa(\alpha - X_{t_i})\Delta). \quad (7.46)$$

Para α

$$\frac{\partial \ln L(\theta)}{\partial \alpha} = \frac{\kappa}{\sigma^2} \sum_{i=1}^n \frac{1}{X_{t_i}} (X_{t_{i+1}} - X_{t_i} - \kappa(\alpha - X_{t_i})\Delta). \quad (7.47)$$

Para σ

$$\frac{\partial \ln L(\theta)}{\partial \sigma} = \frac{1}{\sigma^3} \sum_{i=1}^n \frac{1}{X_{t_i}} (X_{t_{i+1}} - X_{t_i} - \kappa(\alpha - X_{t_i})\Delta)^2 - \frac{n}{\sigma}. \quad (7.48)$$

Los componentes de la matriz (score), es decir (7.46), (7.47) (7.48) son iguales a cero, si y sólo si

$$\sum_{i=1}^n \frac{\alpha - X_{t_i}}{X_{t_i}} (X_{t_{i+1}} - X_{t_i} - \kappa(\alpha - X_{t_i})\Delta) = 0, \quad (7.49)$$

$$\sum_{i=1}^n \frac{1}{X_{t_i}} (X_{t_{i+1}} - X_{t_i} - \kappa(\alpha - X_{t_i})\Delta) = 0, \quad (7.50)$$

$$\sum_{i=1}^n \frac{1}{X_{t_i}} (X_{t_{i+1}} - X_{t_i} - \kappa(\alpha - X_{t_i})\Delta)^2 - n\sigma^2\Delta = 0. \quad (7.51)$$

Resolviendo para σ en la ecuación (7.51)

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n \frac{1}{\bar{X}_i} \left(X_{t_{i+1}} - X_{t_i} - \hat{\kappa}(\hat{\alpha} - X_{t_i})\Delta \right)^2}{\Delta n}}. \quad (7.52)$$

De la ecuación (7.50) se tiene que

$$\hat{\alpha} = \frac{\sum_{i=1}^n \frac{X_{t_{i+1}}}{\bar{X}_i} - n + n\Delta\hat{\kappa}}{\Delta\hat{\kappa}\sum_{i=1}^n \frac{1}{\bar{X}_i}}. \quad (7.53)$$

Tomando la ecuación (7.49) y restando α veces la ecuación (7.50) se obtiene que

$$\sum_{i=1}^n \left(X_{t_{i+1}} - X_{t_i} - \kappa(\alpha - X_{t_i})\Delta \right) = 0. \quad (7.54)$$

Sustituyendo la ecuación resultante (7.53) en la ecuación (7.54) y resolviendo para κ se obtiene que

$$\hat{\kappa} = \frac{n^2 - n\sum_{i=1}^n \frac{X_{t_{i+1}}}{\bar{X}_i} + \sum_{i=1}^n \frac{1}{\bar{X}_i} (X_{t_n} - X_{t_1})}{\Delta \left(n^2 - \sum_{i=1}^n X_{t_i} \sum_{i=1}^n \frac{1}{\bar{X}_i} \right)}. \quad (7.55)$$

7.3. Expansión de la verosimilitud via polinomios de Hermite

Cuando el muestreo se hace a tiempo discreto, los métodos disponibles para obtener la función de verosimilitud se enfocaban en resolver numéricamente la ecuación diferencial parcial de Fokker-Planck también conocida como ecuación forward de Kolmogorov (mencionada en el Capítulo 3) o simular una gran cantidad de trayectorias del proceso. Sin embargo, ningún método produce una expresión cerrada para maximizar sobre θ . Por otro lado, Aït-Sahalia propone una expansión en polinomios de Hermite $p_X^{(J)}$ que aproxima a la probabilidad de transición del proceso $\mathbf{X} = \{X_t\}_{t \geq 0}$ y por lo tanto una sucesión de $\ln^{(J)}$ de aproximaciones de la función de log-verosimilitud. También provee evidencia empírica que $J = 2$ o 3 es ampliamente adecuada para modelos relevantes en finanzas. Puesto que se conoce explícitamente $\ln^{(J)}$, entonces se puede hacer estimación de los parámetros de la misma forma que en el problema de máxima verosimilitud estándar cuando se conoce la función de verosimilitud.

Otra forma de aproximar laprobabilidad de transición de una difusión es expandirla de forma similar a el caso de variables aleatorias independientes e idénticamente distribuidas. La analogía es la siguiente.

Considere una suma estandarizada de variables aleatorias para la cual el teorema del Límite Central se puede aplicar. Para un tamaño de la muestra n suficientemente grande, por el TLC se puede usar la distribución límite $N(0, 1)$, pero si el tamaño de la muestra no es suficientemente grande, se deberá expandir la distribución límite (por ejemplo, por la expansión de Edgerworth) para incrementar la precisión de la aproximación. La idea propuesta por Aït-Sahalia es utilizar dicho enfoque para aproximar la densidad de transición desconocida de una difusión y tratar el caso asintótico cuando $\Delta \rightarrow 0$ en paralelo al caso cuando las variables aleatorias son independientes e idénticamente distribuidas y $n \rightarrow \infty$. El método estándar en el caso que las variables aleatorias son independientes e idénticamente distribuidas fue introducido por Cramér. Sea $p(z)$ la densidad objetivo para expandirse alrededor de la normal estándar $N(0, 1)$. Con este método la expansión converge rápidamente cuando la densidad objetivo es casi normal, o más precisamente cuando sus colas son suficientemente delgadas en el sentido de que $\exp(z^2/2)p'(z)^2$ debe ser integrable. En algunos caso, como en varios modelos de difusión financieros, la suposición Gaussiana no es una opción. Por ejemplo. si el proceso X es un movimiento browniano geométrico, la cola derecha de la correspondiente densidad log-normal $p(z)$ es demasiado larga para converger.

Por lo tanto, Aït-Sahalia propone proceder diferente; en lugar de intentar una expansión directa sobre $p(z)$, primero se transforma la densidad en algo parecido a algo con distribución Normal y luego esta densidad transformada se expande usando polinomios de Hermite.

A lo largo de esta sección se intercambiará la notación de los coeficientes de deriva y de difusión para ser más explícita la dependencia de estas funciones sobre el parámetro θ , como sigue: $\mu(t, X_t) = \mu(X_t; \theta)$ y $\sigma(t; X_t) = \sigma(X_t; \theta)$ debido a que se harán diversos supuestos sobre θ .

Supuestos sobre la primera transformacion

Supongamos que se tiene la ecuación diferencial estocástica

$$dX_t = \mu(X_t; \theta)dt + \sigma(X_t; \theta)dB_t. \quad (7.56)$$

Sea $D_X = (\underline{x}, \bar{x})$ el dominio de la difusión $X_t = \{X_t\}_{t \geq 0}$. Se considerarán los dos caso donde $D_X = (-\infty, +\infty)$ y $D_X = (0, +\infty)$. El último caso es relevante en finanzas, donde se consideran modelos para precios de activos o tasas de interés, donde usualmente μ , y/o σ violan las condiciones de crecimiento lineal cerca de las fronteras o σ es definida

de tal forma que $\lim_{x \rightarrow 0^+} \sigma(x, \theta) = 0$.

Supuesto 7.3.1 (Suavidad de los coeficientes). La función $\mu(x; \theta)$ y $\sigma(x; \theta)$ son infinitamente diferenciables en x y tres veces continuamente diferenciables en θ para todo $x \in D_X$ y $\theta \in \Theta$.

Supuesto 7.3.2 (No degeneración de la difusión). Dos casos pueden suceder:

1. Si $D_X = (-\infty, \infty)$, entonces existe c tal que $\sigma(x; \theta) > c > 0$ para todo $x \in D_X$ y $\theta \in \Theta$.
2. Si $D_X = (0, +\infty)$, entonces $\sigma(x; \theta)$ no se degenera en D_X ; es decir, para cada $\xi > 0$, existe c_ξ ; $\sigma(x, \theta) \geq c_\xi > 0$ para todo $x \in [\xi, \infty]$ y $\theta \in \Theta$. Si además $\lim_{x \rightarrow 0^+} \sigma(x; \theta) = 0$, entonces existe $\xi_0 > 0, \omega > 0, \rho > 0$ tal que $\sigma(x, \theta) \geq \omega x^\rho$ para todo $0 < x \leq \xi_0$ y $\theta \in \Theta$.

El primer paso consiste en estandarizar el coeficiente de difusión del proceso $\mathbf{X} = \{X_t\}_{t \geq 0}$, es decir, transformar $\mathbf{X} = \{X_t\}_{t \geq 0}$ en $\mathbf{Y} = \{Y_t\}_{t \geq 0} = \gamma(\mathbf{X})$, donde $\mathbf{Y} = \{Y_t\}_{t \geq 0}$ satisface la ecuación

$$dY_t = \mu_Y(Y_t, \theta) dt + dB_t \quad (7.57)$$

En el Capítulo 2 se presentó la transformación de Lamperti,

$$Y = \gamma(X, \theta) = \int^X du / \sigma(u; \theta). \quad (7.58)$$

Por la fórmula de Itô, se tiene que

$$dY_t = \mu_Y(Y_t; \theta) dt + dB_t,$$

donde

$$\mu_Y(y; \theta) = \frac{\mu(\gamma^{-1}(y; \theta); \theta)}{\sigma(\gamma^{-1}(y; \theta); \theta)} - \frac{1}{2} \frac{\partial \sigma}{\partial x}(\gamma^{-1}(y; \theta); \theta).$$

Debido a que la transformación de $\mathbf{X} = \{X_t\}_{t \geq 0}$ a $\mathbf{Y} = \{Y_t\}_{t \geq 0}$ puede introducir singularidades en el término de drift, las cuales pueden conducir a propiedades no deseadas, Aït-Sahalia impone algunas restricciones sobre $\mu_Y(y; \theta)$. Las cuales en su mayoría son acerca de las condiciones de crecimiento en las fronteras, de esta forma el proceso $\mathbf{Y} = \{Y_t\}_{t \geq 0}$ se comportar bien para un análisis futuro.

Debido a que $\sigma > 0$ sobre D_X por el supuesto (7.3.2), entonces $\gamma(\mathbf{X})$ es una función creciente e invertible para todo $\theta \in \Theta$.

Se mapea D_X a $D_Y = (\underline{y}, \bar{y})$, el dominio de $\mathbf{Y} = \{Y_t\}_{t \geq 0}$, donde $\underline{y} \equiv \lim_{x \rightarrow \underline{x}^+} \gamma(x; \theta)$ y $\bar{y} \equiv \lim_{x \rightarrow \bar{x}^-} \gamma(x, \theta)$. Por ejemplo, si $D_X = (0, +\infty)$ y $\sigma(x; \theta) = x^\rho$, entonces $\mathbf{X} = (1 - \rho)\mathbf{X}^{1-\rho}$ si $0 < \rho < 1$ (por lo tanto $D_Y = (0, +\infty)$). Si $\rho = 1$ entonces $\mathbf{X} = \ln(\mathbf{X})$ (por lo tanto $D_Y = (-\infty, +\infty)$). Si $\rho < 1$ entonces $\mathbf{Y} = -(\rho - 1)\mathbf{X}^{-(\rho-1)}$ (por lo tanto $D_Y = (-\infty, 0)$).

Supuesto 7.3.3 (Comportamiento en las fronteras). Para todo $\theta \in \Theta$, $\mu_Y(y; \theta)$ y sus derivadas con respecto a y y θ tienen a los mas crecimiento polinomial cerca de las fronteras y $\lim_{y \rightarrow \underline{y}^+} \bar{y}^- \lambda_Y(y; \theta) < +\infty$ donde $-\lambda$ es el potencial, es decir, $\lambda_Y(y; \theta) \equiv -\{\mu_y^2(y; \theta) + \partial \mu_y(y; \theta) / \partial y\} / 2$

1. Frontera izquierda: Si $\underline{y} = 0$, entonces existen constantes $\varepsilon_0, \kappa, \alpha$ tales que para todo $0 < y \leq \varepsilon_0$ y $\theta \in \Theta$, $\mu(y; \theta) \geq \kappa y^{-\alpha}$ donde $\alpha > 1$ y $\kappa > 0$ $\alpha = 1$ y $\kappa \geq 1$. Si $\underline{y} = -\infty$, existen constantes $E_0 > 0$ y $K > 0$ tales que para todo $y \leq -E_0$ y $\theta \in \Theta$, $\mu_y(y; \theta) \geq Ky$.
2. Frontera derecha= Si $\bar{y} = +\infty$, existen constantes $E_0 > 0$ y $K > 0$ tales que para todo $y \geq E_0$ y $\theta \in \Theta$, $\mu_Y(y; \theta) \leq Ky$. Si $\bar{y} = 0$, existen constantes $\varepsilon_0, \kappa, \alpha$ tales que para todo $0 > y \geq -\varepsilon_0$ y $\theta \in \Theta$, $\mu_Y(y; \theta) \leq -\kappa|y|^{-\alpha}$ donde $\alpha > 1$ y $\kappa > 0$ $\alpha = 1$ y $\kappa \geq 1/2$.

Proposición 7.3.4. Bajo los supuestos (7.3.1), (7.3.2), (7.3.3), la ecuación diferencial estocástica (7.58) admite una solución débil $\mathbf{Y} = \{Y_t\}_{t \geq 0}$ única en ley de probabilidad para cada distribución con valor inicial Y_0 . Las fronteras de D_Y son inalcanzables en el sentido de que $\mathbb{P}(T_Y = +\infty) = 1$, donde $T_Y \equiv \inf\{t \geq 0 : Y_t \notin D_Y\}$.

Demostración. Ver [Aït-Sahalia 2002].

Proposición 7.3.5. Bajo los supuestos (7.3.1), (7.3.2), (7.3.3), $\mathbf{Y} = \{Y_t\}_{t \geq 0}$ admite una probabilidad de transición $p_Y(\Delta, y|y_0, \theta)$, que es continuamente diferenciable en $\Delta > 0$, infinitamente diferenciable en $y \in D_Y$ y $y_0 \in D_Y$, y tres veces continuamente diferenciable en $\theta \in \Theta$. Mas aún, existe $\Delta > 0$ tal que para cada $\Delta \in (0, \bar{\Delta})$, existen constantes positivas $C_i, i = 0, \dots, 4$ y D_0 tales que para cada $\theta \in \Theta$ y $(y, y_0) \in D_Y$

1. $0 < p_Y(\Delta, y|y_0; \theta) \leq C_0 \Delta^{-1/2} e^{-3(y-y_0)^2/8\Delta} e^{C_1|y-y_0||y_0| + C_2|y-y_0| + C_3|y_0| + C_4 y_0^2}$,
2. $|\partial p_Y(\Delta, y|y_0; \theta) \partial y| \leq D_0 \Delta^{-1/2} e^{-3(y-y_0)^2/8\Delta} P(|y|, |y_0|) e^{C_1|y-y_0||y_0| + C_2|y-y_0| + C_3|y_0| + C_4 y_0^2}$,

donde P es un polinomio de orden finito en y_0 y y .

Demostración. Ver [Aït-Sahalia 2002].

La cota de la proposición anterior para p_Y implica que las colas de p_Y tienen un comportamiento gaussiano debido al decaimiento exponencial. Aït-Sahalia indica que este es un requisito esencial para que la expansión en polinomios de Hermite converja. Sin embargo, aunque la probabilidad de transición del proceso $\mathbf{Y} = \{Y_t\}_{t \geq 0}$ es más “cercana” a una variable aleatoria Normal que la transición de $\mathbf{X} = \{X_t\}_{t \geq 0}$, no es práctico expandir p_Y . Esto debido a que la probabilidad de transición p_Y se parece cada vez más a la delta de Dirac cuando $\Delta \rightarrow 0$. Por esta razón Aït-Sahalia sugiere una segunda transformación al proceso $\mathbf{X} = \{X_t\}_{t \geq 0}$.

El siguiente resultado muestra que éstas propiedades se extienden esencialmente a la difusión $\mathbf{X} = \{X_t\}_{t \geq 0}$ del interés original.

Proposición 7.3.6. Bajo los supuestos (7.3.1), (7.3.2), 7.3.3), la ecuación diferencial estocástica (7.56) admite una solución débil $\{X_t\}_{t \geq 0}$, única en ley de probabilidad, para cada distribución con valor inicial X_0 . Las fronteras de D_X son inalcanzables, en el sentido de que $\mathbb{P}(T_X = \infty) = 1$ donde $T_X \equiv \inf\{t \geq 0 | X_t \notin D_X\}$. Además, $\{X_t\}_{t \geq 0}$ admite una probabilidad de transición $p_X(\Delta, x | x_0; \theta)$ la cual es continuamente diferenciable en $\Delta > 0$, infinitamente diferenciable en $x \in D_X$ y $X_0 \in D_X$ y tres veces continuamente diferenciable en $\theta \in \Theta$.

Segunda transformación

Haciendo la transformación de $\mathbf{X} = \{X_t\}_{t \geq 0}$ a $\mathbf{Y} = \{Y_t\}_{t \geq 0}$ es posible construir una expansión de la densidad de transición para $\mathbf{Y} = \{Y_t\}_{t \geq 0}$. Sin embargo, puesto que la probabilidad de transición de la cual se tiene interés es la de $\mathbf{X} = \{X_t\}_{t \geq 0}$, es necesario encontrar la relación entre las dos probabilidades de transición, la cual se puede encontrar usando la fórmula del Jacobiano, es decir,

$$\begin{aligned}
 p_X(\Delta, x | x_0; \theta) &= \frac{\partial}{\partial x} P(X_{t+\Delta} \leq x | X_t = x_0; \theta) \\
 &= \frac{\partial}{\partial x} P(Y_{t+\Delta} \leq \gamma(x; \theta) | Y_t = \gamma(x_0; \theta); \theta) \\
 &= \frac{\partial}{\partial x} \left(\int^{\gamma(x; \theta)} p_Y(\Delta, y | \gamma(y_0; \theta)) dy \right) \\
 &= \frac{p_Y(\Delta, \gamma(x; \theta) | \gamma(x_0; \theta); \theta)}{\sigma(\gamma(x; \theta); \theta)}.
 \end{aligned} \tag{7.59}$$

El proceso transformado $\mathbf{Y} = \{Y_t\}_{t \geq 0}$ satisface la siguiente ecuación diferencial estocástica

$$dY_t = \mu_Y(Y_t; \theta)dt + dB_t.$$

Para $\Delta > 0$, $\theta \in \Theta$, y $y_0 \in \mathbb{R}$, se define el incremento “pseudo-normalizado” de $\mathbf{Y} = \{Y_t\}_{t \geq 0}$ como

$$Z_t \equiv \Delta^{-1/2}(Y_t - y_0). \quad (7.60)$$

Aït-Sahalia muestra que para Δ fija, Z_t definida como en (7.60) resulta ser lo suficientemente cercana a una variable $N(0, 1)$ para hacer posible una expansión en series convergente a p_Z alrededor de una distribución $N(0, 1)$.

Sea $p_Y(\Delta, y|y_0; \theta)$ la densidad condicional de $Y_{t+\Delta}|Y_t$, y se define la función de densidad de Z_t como

$$p_Z(\Delta, z|y_0; \theta) \equiv \Delta^{1/2} p_Y(\Delta, \Delta^{1/2}z + y_0|y_0; \theta). \quad (7.61)$$

Una vez que se obtiene una sucesión de aproximaciones para la función $(z, y_0) \mapsto p_Y(\Delta, y|y_0; \theta)$, se infiere una sucesión de aproximaciones a la función $(y, y_0) \mapsto p_Y(\Delta, y|y_0; \theta)$ invirtiendo la probabilidad de transición (7.61)

$$p_Y(\Delta, y|y_0; \theta) \equiv \Delta^{-1/2} p_Z(\Delta, \Delta^{-1/2}(y - y_0)|y_0; \theta), \quad (7.62)$$

y entonces como se mencionó anteriormente se puede encontrar la densidad de $\mathbf{X} = \{X_t\}_{t \geq 0}$ por la fórmula del Jacobiano para el cambio de probabilidad de transición como sigue:

$$p_X(\Delta, x|x_0; \theta) = \sigma(x, \theta)^{-1} \times p_Y(\Delta, \gamma(x; \theta)|\gamma(x_0; \theta); \theta). \quad (7.63)$$

7.3.1. Aproximación de la función de transición por polinomios de Hermite

Los polinomios de Hermite clásicos son

$$H_j(z) \equiv e^{z^2/2} \frac{d^j}{dz^j} [e^{-z^2/2}], \quad j \geq 0 \quad (7.64)$$

y sea $\phi(z)$ la función de densidad de una variable aleatoria $N(0, 1)$, es decir, $\phi(z) \equiv \frac{e^{-z^2/2}}{\sqrt{2\pi}}$

Los primeros 7 polinomios de Hermite son:

$$\begin{aligned} H_0(z) &= 1, \\ H_1(z) &= -z, \\ H_2(z) &= z^2 - 1, \\ H_3(z) &= -z^3 + 3z, \\ H_4(z) &= z^4 - 6z^2 + 3, \\ H_5(z) &= -z^5 + 10z^3 - 15z, \\ H_6(z) &= z^6 - 15z^4 + 45z^2 - 15. \end{aligned}$$

A continuación se enunciará una propiedad importante de los polinomios de Hermite.

7.3.1.1. Ortogonalidad

Sea $H_n(z)$ un polinomio de Hermite de grado n . Estos polinomios son ortogonales con respecto a la función de peso $w(z) = e^{-\frac{z^2}{2}}$, es decir,

$$\int_{-\infty}^{\infty} H_m(z) H_n(z) w(z) dz = 0, \quad m \neq n.$$

Más aún,

$$\int_{-\infty}^{\infty} H_m(z) H_n(z) e^{-\frac{z^2}{2}} dz = \sqrt{2\pi n!} \delta_{nm}.$$

donde $\delta_{nm} = 1$ si $m = n$ y $\delta_{nm} = 0$ si $m \neq n$.

Puesto que la probabilidad de transición del proceso Z_t se parece bastante a la densidad de una variable aleatoria $N(0, 1)$ entonces es posible expandir la densidad del proceso Z_t en términos de polinomios de Hermite, a través de la siguiente fórmula

$$p_Z^{(J)}(\Delta, z|y_0; \theta) = \phi(z) \sum_{j=0}^J \eta_Z^{(j)}(\Delta, y_0; \theta) H_j(z). \quad (7.65)$$

donde $\phi(Z)$ es la densidad de una distribución $N(0, 1)$.

La razón por la cual se usan polinomios de Hermite es debido a que son ortogonales con respecto al producto escalar ponderado por la densidad Normal, $L^2(\phi)$. Debido a la propiedad ortonormal de estos polinomios, entonces cualquier función en $L^2(\phi)$ puede ser expresada como combinación lineal de estos polinomios.

Los coeficientes en $\eta_Z^{(j)}$ pueden ser encontrados multiplicando ambos lados de (7.65) por $H_j(z)$, integrando de $-\infty$ a ∞ , cambiando el orden de la integral y la suma y usando la propiedad ortogonal, es decir,

$$\int_{-\infty}^{\infty} p_Z^{(j)}(\Delta, z|y_0; \theta) H_j(z) dz = \int_{-\infty}^{\infty} \phi(z) \sum_{j=0}^J \eta_Z^{(j)}(\Delta, y_0; \theta) H_j(z) H_j(z) dz = \eta_Z^{(j)} j!. \quad (7.66)$$

Por lo tanto,

$$\eta_Z^{(j)}(\Delta, y_0; \theta) = \frac{1}{j!} \int_{-\infty}^{\infty} p_Z^{(j)}(\Delta, z|y_0; \theta) H_j(z) dz. \quad (7.67)$$

Por la fórmula del Jacobiano mencionada anteriormente

$$\eta_Z^{(j)}(\Delta, y_0; \theta) = \frac{1}{j!} \int_{-\infty}^{\infty} p_Z^{(j)}(\Delta, z|y_0; \theta) H_j(z) dz \quad (7.68)$$

$$= \frac{1}{j!} \int_{-\infty}^{\infty} H_j(z) \Delta^{1/2} p_Y(\Delta, \Delta^{1/2} z + y_0|y_0; \theta) dz \quad (7.69)$$

$$= \frac{1}{j!} \int_{-\infty}^{\infty} H_j(\Delta^{-1/2}(y - y_0)) p(Y)(\Delta, y|y_0; \theta) dy \quad (7.70)$$

$$= \frac{1}{j!} \mathbb{E}[H_j(\Delta^{-1/2}(Y_{t+\Delta} - y_0)) | Y_t = y_0; \theta]. \quad (7.71)$$

Por lo tanto, los coeficientes $\eta_Z^{(j)}$ son momentos condicionados del proceso $Y_t = \{Y_t\}_{t \geq 0}$. Existen varias formas de calcular estos coeficientes. Una forma de resolver esta esperanza es calcular explícitamente una expansión en series de Taylor en Δ para los coeficientes $\eta_Z^{(j)}$. Ait-Sahalia demostró que para una función suave g y un entero n , se tiene que

$$\mathbb{E}[g(Y_{t+\Delta}) | Y_t = y_0] = \sum_{i=1}^n (\mathcal{L}_\theta^i g)(y_0) \frac{\Delta^i}{i!} + \text{residuo}, \quad (7.72)$$

donde \mathcal{L}_θ es el generador infinitesimal de la difusión $Y_t = \{Y_t\}_{t \geq 0}$, definido como

$$(\mathcal{L}_\theta g)(y_0) = \mu_Y(y_0, \theta) g'(y_0) + \frac{1}{2} g''(y_0). \quad (7.73)$$

La notación $(\mathcal{L}_\theta^i g)$ significa que el generador es aplicado recursivamente i veces. Por ejemplo:

$$\begin{aligned}\mathcal{L}_\theta^0 g &= g, \\ \mathcal{L}_\theta^1 g &= \mu_Y g^{(1)} + \frac{1}{2} g^{(2)}, \\ \mathcal{L}_\theta^2 g &= \mu_Y \left(\mu_Y g^{(2)} + \mu_Y^{(1)} + \frac{1}{2} g^{(3)} \right) + \frac{1}{2} \left(\mu_Y^{(1)} + \mu_Y g^{(3)} + \mu_Y^{(2)} g^{(1)} + \mu_Y^{(1)} g^{(2)} + \frac{1}{2} g^{(4)} \right) \\ &= \mu^2 g^{(2)} + \mu_Y \mu_Y^{(1)} g^{(1)} + \mu_Y g^{(3)} + \frac{1}{2} \mu_Y^{(2)} g^{(1)} + \mu_Y^{(1)} g^{(2)} + \frac{1}{4} g^{(4)}.\end{aligned}$$

donde $\mu_Y^{(k)}$ y $g^{(k)}$ representan las k -ésimas derivadas de las funciones μ y g respectivamente.

Por lo tanto, los coeficientes $\eta_Z^{(j)}$ se pueden calcular de la siguiente forma

$$\eta_Z^{(j)} = \frac{1}{j!} \sum_{i=1}^n (\mathcal{L}_\theta^i [H_j(z)])(y_0) \frac{\Delta^i}{i!} + \text{residuo}.$$

Por ejemplo para $H_1(z) = -z$, se tiene que

$$\begin{aligned}\mathcal{L}_\theta^0 [H_1(z)] &= \frac{-\mu}{\sqrt{\Delta}}, \\ \mathcal{L}_\theta^1 [H_1(z)] &= \frac{-2\mu\mu^{(1)} + \mu^{(2)}}{2\sqrt{\Delta}}, \\ \mathcal{L}_\theta^2 [H_1(z)] &= \frac{4\mu^2\mu^{(2)} + 4\mu(\mu^{(1)})^2 + 4\mu\mu^{(2)} + 6\mu^{(1)}\mu^{(2)} + \mu^{(4)}}{4\sqrt{\Delta}}.\end{aligned}$$

Estos tres componentes son suficientes para calcular el coeficiente $\eta_Z^{(1)}$ hasta $n = 3$, a saber,

$$\begin{aligned}\eta_Z^1 &= -\mu\Delta^{1/2} - (2\mu\mu^{(1)} + \mu^{(2)})\Delta^{3/2}/4 \\ &\quad - (4\mu(\mu^{(1)})^2 + 4\mu^2\mu^{(2)} + 6\mu^{(1)}\mu^{(2)} + 4\mu\mu^{(3)} + \mu^{(4)})\Delta^{5/2}/24.\end{aligned}$$

Análogamente, se pueden calcular los coeficientes $\eta_Z^{(j)}$ para los primeros 7 polinomios de Hermite. Aït-Sahalia usó este método para calcular explícitamente los primeros 6

coeficientes hasta $n = 3$. Los coeficientes son:

$$\begin{aligned}
 \eta_Z^0 &= 1 \\
 \eta_Z^1 &= -\mu\Delta^{1/2} - (2\mu\mu^{(1)} + \mu^{(2)})\Delta^{3/2}/4 \\
 &\quad - (4\mu(\mu^{(1)})^2 + 4\mu^2\mu^{(2)} + 6\mu^{(1)}\mu^{(2)} + 4\mu\mu^{(3)} + \mu^{(4)})\Delta^{5/2}/24, \\
 \eta_Z^2 &= (\mu^2 + \mu^{(1)})\Delta/2 + (6\mu^2\mu^{(1)} + 4(\mu^{(1)})^2 + 7\mu\mu^{(2)} + 2\mu^{(3)})\Delta^2/12 \\
 &\quad + (28\mu^2(\mu^{(1)})^2 + 28\mu^2\mu^{(3)} + 16(\mu^{(1)})^3 + 16\mu^3\mu^{(2)} + 88\mu\mu^{(1)}\mu^{(2)} \\
 &\quad + 21(\mu^{(2)})^2 + 32\mu^{(1)}\mu^{(3)} + 16\mu\mu^{(4)} + 3\mu^{(5)})\Delta^3/96, \\
 \eta_Z^3 &= -(\mu^3 + 3\mu\mu^{(1)} + \mu^{(2)})\Delta^{3/2}/6 - (12\mu^3\mu^{(1)} + 28\mu(\mu^{(1)})^2 \\
 &\quad + 22\mu^2\mu^{(2)} + 24\mu^{(1)}\mu^{(2)} + 14\mu\mu^{(3)} + 3\mu^{(4)})\Delta^{5/2}/48, \\
 \eta_Z^4 &= (\mu^4 + 6\mu^2\mu^{(1)} + 3(\mu^{(1)})^2 + 4\mu\mu^{(2)} + \mu^{(3)})\Delta^2/24 \\
 &\quad + (20\mu^4\mu^{(1)} + 50\mu^3\mu^{(2)} + 100\mu^2(\mu^{(1)})^2 + 50\mu^2\mu^{(3)} + 23\mu\mu^{(4)} \\
 &\quad + 180\mu\mu^{(1)}\mu^{(2)} + 40(\mu^{(1)})^3 + (34\mu^{(2)})^2 + 52\mu^{(1)}\mu^{(3)} + 4\mu^{(5)})\Delta^3/240, \\
 \eta_Z^5 &= -(\mu^5 + 10\mu^3\mu^{(1)} + 15\mu(\mu^{(1)})^2 + 10\mu^2\mu^{(2)} \\
 &\quad + 10\mu^{(1)}\mu^{(2)} + 5\mu\mu^{(3)} + \mu^{(4)})\Delta^{5/2}/120, \\
 \eta_Z^6 &= (\mu^6 + 15\mu^4\mu^{(1)} + (15\mu^{(1)})^3 + 20\mu^3\mu^{(2)} + 15\mu^{(1)}\mu^{(3)} + 45\mu^2(\mu^{(1)})^2 \\
 &\quad + 10(\mu^{(2)})^2 + 15\mu^2\mu^{(3)} + 60\mu\mu^{(1)}\mu^{(2)} + 6\mu\mu^{(4)} + \mu^{(5)})\Delta^3/720,
 \end{aligned}$$

Haciendo una analogía a la construcción de la ecuación (7.62). Aït-Sahalia construye una sucesión de aproximaciones para p_Y como sigue

$$p_Y^{(J)}(\Delta, y|y_0; \theta) \equiv \Delta^{-1/2} p_Z^{(J)}(\Delta, \Delta^{-1/2}(y - y_0)|y_0; \theta), \quad (7.74)$$

y finalmente aproxima p_X imitando la construcción de la ecuación (7.63), es decir,

$$p_X^{(J)}(\Delta, x|x_0; \theta) \equiv \sigma(x; \theta)^{-1} p_Y^{(J)}(\Delta, \gamma(x, \theta)|\gamma(x_0; \theta); \theta). \quad (7.75)$$

El siguiente teorema demuestra que la expresión (7.75) converge uniformemente cuanto más términos sean añadidos, y que el límite es de hecho la verdadera (pero desconocida) función de densidad p_X .

Teorema 7.3.7. Bajo los supuestos (7.3.1),(7.3.2),(7.3.3), existe $\bar{\Delta} > 0$ (dado en la Proposición 7.3.4) tal que para cada $\Delta \in (0, \bar{\Delta})$, $\theta \in \Theta$ y $(x, x_0) \in D_X^2$:

$$p_X^{(J)}(\Delta, x|x_0; \theta) \xrightarrow{J \rightarrow \infty} p_X(\Delta, x|x_0; \theta). \quad (7.76)$$

Demostración. Ver [Aït-Sahalia 2002].

7.3.1.2. Expansión en polinomios de Hermite para el proceso de Ornstein-Uhlenbeck

Consideremos la ecuación diferencial estocástica definida como:

$$dX_t = \theta(\mu - X_t)dt + \sigma dB_t. \quad (7.77)$$

El primer paso para aproximar la probabilidad de transición es hacer el término de difusión unitario en la ecuación diferencial estocástica (7.77).

Aplicando la transformación de Lamperti se tiene que el nuevo proceso \mathbf{Y} tiene la siguiente forma

$$\mathbf{Y} = \gamma(\mathbf{X}, \theta) = \int^{x_t} \frac{1}{\sigma} du = \frac{\mathbf{X}}{\sigma}.$$

Utilizando la fórmula de Itô se obtiene que

$$dY_t = \left(\frac{\theta\mu}{\sigma} - \theta Y_t \right) dt + dB_t,$$

de la ecuación anterior obtenemos que

$$\mu_Y(y; \theta) = \frac{\theta\mu}{\sigma} - \frac{\theta}{y}, \quad (7.78)$$

por lo que $\mu^{(1)} = -\theta$ y todas las derivadas de orden mayor son cero.

7.3.1.3. Expansión en polinomios de Hermite para el modelo de Cox-Ingersoll-Ross

Consideremos la ecuación diferencial estocástica definida como:

$$dX_t = (\kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dB_t), \quad X_0 = x_0 > 0, \quad (7.79)$$

El primer paso para aproximar la densidad es hacer el término de difusión unitario en la ecuación diferencial estocástica (7.79).

Aplicando la transformación de Lamperti se tiene que el nuevo proceso \mathbf{Y} tiene la siguiente forma

$$\mathbf{Y} = \gamma(\mathbf{X}, \theta) = \int^{x_t} \frac{1}{\sigma\sqrt{u}} du = \frac{2\sqrt{\mathbf{X}}}{\sigma}$$

Utilizando la fórmula de Itô se obtiene que

$$dY_t = \left(\left(\frac{2\kappa\alpha}{\sigma^2} - \frac{1}{2} \right) \frac{1}{Y_t} - \frac{\kappa Y_t}{2} \right) dt + dB_t,$$

de la ecuación anterior obtenemos que

$$\begin{aligned}\mu_Y(y, \theta) &= \left(\frac{2\kappa\alpha}{\sigma^2} - \frac{1}{2} \right) \frac{1}{Y_t} - \frac{\kappa Y_t}{2} \\ &= \frac{4\kappa\alpha - \sigma^2}{2\sigma y} - \frac{\kappa y}{2}.\end{aligned}$$

Entonces, es posible calcular las derivadas para μ_Y , a saber,

$$\begin{aligned}\mu^{(1)}(y, \theta) &= -\frac{\kappa}{2} - \frac{4\kappa\alpha - \sigma^2}{2y^2\sigma^2}, \\ \mu^{(2)}(y, \theta) &= \frac{4\kappa\alpha - \sigma^2}{y^3\sigma^2}, \\ \mu^{(3)}(y, \theta) &= -\frac{3(4\kappa\alpha - \sigma^2)}{y^4\sigma^2}, \\ \mu^{(4)}(y, \theta) &= \frac{12(4\kappa\alpha - \sigma^2)}{y^5\sigma^2}, \\ \mu^{(5)}(y, \theta) &= -\frac{60(4\kappa\alpha - \sigma^2)}{y^6\sigma^2}, \\ \mu^{(6)}(y, \theta) &= \frac{360(4\kappa\alpha - \sigma^2)}{y^7\sigma^2}.\end{aligned}$$

7.3.1.4. Expansión en polinomios de Hermite para el modelo de Black-Scholes-Merton

Consideremos la ecuación diferencial estocástica definida como:

$$dX_t = \mu X_t dt + \sigma X_t dB_t. \quad (7.80)$$

El primer paso para aproximar la densidad es hacer el término de difusión unitario en la ecuación diferencial estocástica (7.80).

Aplicando la transformación de Lamperti se tiene que el nuevo proceso Y tiene la siguiente forma

$$Y = \gamma(\mathbf{X}, \theta) = \int^{x_t} \frac{1}{\sigma u} du = \frac{\log(\mathbf{X})}{\sigma}.$$

Utilizando la fórmula de Itô se obtiene que

$$dY_t = \left(\frac{\mu}{\sigma} - \frac{\sigma}{2} \right) dt + dB_t,$$

de la ecuación anterior obtenemos que todas las derivadas de μ_Y son cero.

Capítulo 8

Conclusiones

El propósito de este trabajo fue revisar algunos de los métodos numéricos que se tienen para simular ecuaciones diferenciales estocásticas que toman valores en \mathbb{R} .

Se revisaron tres de los procesos de difusión más conocidos y utilizados en el mundo financiero, así como sus principales propiedades. El análisis de estos procesos en general, así como sus propiedades más importantes, la relación que tienen con las ecuaciones diferenciales estocásticas y con los procesos de Markov se revisó en el Capítulo 4.

El capítulo más interesante de este trabajo es el Capítulo 7, en el cual se revisaron tres métodos de estimación. Para el método de estimación por máxima verosimilitud se hizo hincapié que la pieza fundamental es calcular la probabilidad de transición del proceso y poder hacer los cálculos analíticos para obtener los estimadores. Este método es el más efectivo, sin embargo, la probabilidad de transición se conoce para muy pocos procesos de difusión y aún conociendo la probabilidad de transición del proceso, en muchas ocasiones llega a ser difícil desde el punto de vista analítico maximizar la verosimilitud, como es el caso del proceso CIR. El segundo método que se revisó fue el método de estimación máxima verosímil vía la discretización de Euler-Maruyama. A pesar de que este método en teoría es sencillo de implementar, puesto que se supone que la distribución condicional del proceso es Normal, el método llega a ser muy poco efectivo si la distancia entre las observaciones no es pequeña, debido al supuesto de normalidad. La ventaja de este método es que se puede usar fácilmente para cualquier difusión únicamente identificando los coeficientes de deriva y de difusión. El tercer método consiste en aproximar a la probabilidad de transición por una expansión en series. Este método se puede aplicar a cualquier difusión tal que sus coeficientes cumplan las condiciones impuestas. El costo computacional de este método puede llegar a ser muy caro y está en duda el orden de expansión que propone el autor. Se puede concluir que el método más preciso para estimación de parámetros es el método de máxima verosimilitud.

Al final se agregó un apéndice de simulación en R para la estimación de los parámetros para el procesos de Ornstein-Uhlenbeck simulado con la pequetería sde .

Comparaciones

Para finalizar este capítulo se comparará la eficacia de los estimadores (máximo verosímil y Euler-Maruyama) para el modelo de Ornstein-Uhlenbeck variando Δ .

Supongamos que tenemos la siguiente ecuación diferencial estocástica

$$dX_t = 10(8 - X_t)dt + 10dB_t.$$

$\Delta = 0.5$

```
Estimacion_MV_Vasicek(50000, 10, 8, 10, .5)
```

```
##      theta1    theta2    theta3
## 1  9.397503  7.489747  9.653991
```

```
Estimacion_EM_Ornstein(50000, 10, 8, 10, .5, 10/8)
```

```
##      theta1    theta2    theta3
## 1  2.435501  1.954805  3.898383
```

$\Delta = 0.05$

```
Estimacion_MV_Vasicek(50000, 10, 8, 10, .05)
```

```
##      theta1    theta2    theta3
## 1  9.93561  8.043405  7.470886
```

```
Estimacion_EM_Ornstein(50000, 10, 8, 10, .05, 10/8)
```

```
##      theta1    theta2    theta3
## 1  8.514834  6.597143  13.505
```

$\Delta = 0.005$

```
Estimacion_MV_Vasicek(50000, 10, 8, 10, .005)
```

```
##      theta1    theta2    theta3
## 1  9.132045  7.734093  2.708827
```

```
Estimacion_EM_Ornstein(50000, 10, 8, 10, .005, 10/8)
```

```
##      theta1    theta2    theta3
## 1  9.925583  8.115821  10.97455
```

$\Delta = 0.0025$

```
Estimacion_MV_Vasicek(50000, 10, 8, 10, .0025)
```

```
##      theta1    theta2    theta3  
## 1 10.75429  7.617884  1.922395
```

```
Estimacion_EM_Ornstein(50000, 10, 8, 10, .0025, 10/8)
```

```
##      theta1    theta2    theta3  
## 1 13.20602  8.398276 11.05083
```

Se observa que los estimadores más precisos cuando $\Delta = 0.5$ y $\Delta = 0.005$ se obtienen por el método de máxima verosimilitud. Por otro lado cuando $\Delta = 0.05$ y $\Delta = 0.025$, el estimador por el método de Euler-Maruyama, es más preciso, como establece la teoría.

Apéndice A

Simulación en R

A.1. Estimación máxima verosímil

A.1.1. Estimación máxima verosímil para el proceso de Ornstein-Uhlenbeck

Descripción

Estimacion_MV_Vasicek() hace uso de la paquetería “sde”, simula una trayectoria del proceso de Ornstein-Uhlenbeck y estima los parámetros mediante el algoritmo descrito en la subsección 7.1.1.

```
library(sde)
Estimacion_MV_Vasicek<-function(N, theta1, theta2, theta3,
                                deltat){
  X <- sde.sim(model="OU", X0=theta1/theta2,
               theta=c(theta1,theta2,theta3), N=N, delta=deltat)
  n <- length(X)
  tmp.sum <- sum(X[1:(n-1)]*X[2:n])
  dt <- deltat(X)

  j=(n*sum(X[2:n]*X[1:(n-1)])-sum(X[1:(n-1)])*sum(X[2:n]))/
    (n*(sum(X[2:n]^2))-sum(X[2:n])^2)
  theta2.gorro=-log(j)/dt;

  k=(sum(X[2:n])-j*sum(X[1:(n-1)]))/n;
  theta1.gorro=k/(1-j)*theta2.gorro; theta1.gorro

  h=(sum((X[2:n]-j*X[1:(n-1)]-k)^2))/n
  theta3.gorro=sqrt((h*2*theta2.gorro)/(1-exp(-theta2.gorro)))

  estimadores=data.frame(theta1.gorro, theta2.gorro,theta3.gorro)
```

```

colnames(estimadores) <- c("theta1", "theta2", "theta3")
return(estimadores)
}

```

```
Estimacion_MV_Vasicek(100, 10, 6, 4, .05)
```

```

##
## T set to = 5.000000
##      theta1   theta2   theta3
## 1 12.54014  6.784075  2.566748

```

Argumentos

- **N** Número de pasos a simular.
- **theta1** parámetro del proceso.
- **theta2** parámetro del proceso.
- **theta3** parámetro del proceso.
- **deltat** Tiempo entre cada paso de la simulación.

A.1.2. Estimación máxima verosímil para el modelo de Black-Scholes-Merton

Descripción

Estimacion_MV_BSM() hace uso de la paqueteríaa “sde”, simula una trayectoria del movimiento browniano geométrico y estima los parámetros mediante el algoritmo descrito en la subsección 7.1.2

Código

```

library(sde)

Estimacion_MV_BSM <- function(N, theta1, theta2, deltat, X0){
  X <- sde.sim(model="BS", theta=c(theta1, theta2), N=N, delta=deltat,
              X0=X0)
  n <- length(X)
  mu.gorro <- (log(X[n]) - log(X[1])) / (deltat * n)
  sigma.gorro <- sum((log(X[2:n]) - log(X[1:(n-1)])) -
                    mu.gorro * deltat)^2) / (n * deltat)
}

```

```
estimadores<-data.frame(mu.gorro, sqrt(sigma.gorro))
colnames(estimadores)<-c("m", "sigma")
return(estimadores)
}
```

Argumentos

- **N** Número de pasos a simular.
- **theta1** parámetro de deriva del proceso.
- **theta2** parámetro de difusión del proceso.
- **deltat** tiempo entre cada paso de la simulación.
- **X0** valor inicial del proceso.

A.1.3. Estimación máxima verosímil para n trayectorias del proceso de Ornstein-Uhlenbeck

Descripción

Vasicek_media_varianza_intervalos() simula $n1$ trayectorias del proceso de Ornstein-Uhlenbeck, mediante la función `sde.sim()` de la paquetería “sde”, hace uso de la función `Estimacion_MV_Vasicek()` para estimar los parámetros $(\theta_1, \theta_2, \theta_3)$ de las $n1$ trayectorias, promedia las $n1$ estimaciones para cada parámetro, calcula la media muestral, varianza muestral y construye intervalos de confianza del 95 % para cada parámetro. Toma cuatro submuestras de $n1$ de tamaño $n2, n3, n4$ e implementa los mismos pasos mencionados anteriormente para cada submuestra.

Código

```
library(sde)

## función para calcular la varianza por renglones
Varianza_renglones<- function(x) { #
  rowSums((x - rowMeans(x))^2)/(dim(x)[2] - 1)
}

Vasicek_media_varianza_intervalos<-function(N,theta1, theta2, theta3,
  delta, n1 ,n2 ,n3, n4){

  X1<-replicate(n1*2,Estimacion_MV_Vasicek(N, theta1, theta2,
  theta3,delta))
  X1=t(na.omit(t(X1)))
  X<-t(head(t(X1),n1))
  I1=quantile(X[1,], probs = c(0.25,0.975), na.rm=TRUE)
  I2=quantile(X[2,], probs = c(0.25,0.975), na.rm=TRUE)
  I3=quantile(X[3,], probs = c(0.25,0.975), na.rm=TRUE)

  M1=colMeans(X,na.rm = TRUE)
  M2=Varianza_renglones(X)
  Theta1_n1<-matrix(c(colMeans(t(X))[1], Varianza_renglones(X)[1],
  I1),1,4)
  dimnames(Theta1_n1)=list(c(n1),
  c("Media","Varianza", "25%", "97.5%"))

  Theta2_n1<-matrix(c(colMeans(t(X))[2], Varianza_renglones(X)[2],
  I2),1,4)
  dimnames(Theta2_n1)=list(n1,
  c("Media","Varianza", "25%", "97.5%"))

  Theta3_n1<-matrix(c(colMeans(t(X))[3], Varianza_renglones(X)[3],
```

```

I3), 1, 4)
dimnames(Theta3_n1)=list(c(n1),
                          c("Media", "Varianza", "25%", "97.5%"))

X2<-t(head(t(X1), n2))
I1_n2=quantile(X2[1, ], probs = c(0.25, 0.975), na.rm=TRUE)
I2_n2=quantile(X2[2, ], probs = c(0.25, 0.975), na.rm=TRUE)
I3_n2=quantile(X2[3, ], probs = c(0.25, 0.975), na.rm=TRUE)

M1=colMeans(X2, na.rm = TRUE)
M2=Varianza_renglones(X2)
Theta1_n2<-matrix(c(colMeans(t(X2)) [1], Varianza_renglones(X2) [1],
I1_n2), 1, 4)
dimnames(Theta1_n2)=list(c(n2),
                          c("Media", "Varianza", "25%", "97.5%"))

Theta2_n2<-matrix(c(colMeans(t(X2)) [2], Varianza_renglones(X2) [2],
I2_n2), 1, 4)
dimnames(Theta2_n2)=list(c(n2),
                          c("Media", "Varianza", "25%", "97.5%"))

Theta3_n2<-matrix(c(colMeans(t(X2)) [3], Varianza_renglones(X2) [3],
I3_n2), 1, 4)
dimnames(Theta3_n2)=list(c(n2),
                          c("Media", "Varianza", "25%", "97.5%"))

X3<-t(head(t(X1), n3))
I1_n3=quantile(X3[1, ], probs = c(0.25, 0.975), na.rm=TRUE)
I2_n3=quantile(X3[2, ], probs = c(0.25, 0.975), na.rm=TRUE)
I3_n3=quantile(X3[3, ], probs = c(0.25, 0.975), na.rm=TRUE)

M1=colMeans(X3, na.rm = TRUE)
M2=Varianza_renglones(X3)
Theta1_n3<-matrix(c(colMeans(t(X3)) [1], Varianza_renglones(X3) [1],
I1_n3), 1, 4)
dimnames(Theta1_n3)=list(c(n3),
                          c("Media", "Varianza", "25%", "97.5%"))

Theta2_n3<-matrix(c(colMeans(t(X3)) [2], Varianza_renglones(X3) [2],
I2_n3), 1, 4)
dimnames(Theta2_n3)=list(c(n3),
                          c("Media", "Varianza", "25%", "97.5%"))

Theta3_n3<-matrix(c(colMeans(t(X3)) [3], Varianza_renglones(X3) [3],
I3_n3), 1, 4)
dimnames(Theta3_n3)=list(c(n3),
                          c("Media", "Varianza", "25%", "97.5%"))

```



```

X4<-t(head(t(X1),n4))
I1_n4=quantile(X4[1,], probs = c(0.25,0.975),na.rm=TRUE)
I2_n4=quantile(X4[2,], probs = c(0.25,0.975),na.rm=TRUE)
I3_n4=quantile(X4[3,], probs = c(0.25,0.975),na.rm=TRUE)

M1=colMeans(X4,na.rm = TRUE)
M2=Varianza_renglones(X4)
Theta1_n4<-matrix(c(colMeans(t(X4))[1], Varianza_renglones(X4)[1],
I1_n4), 1, 4)
dimnames(Theta1_n4)=list(c(n4),
                          c("Media","Varianza", "25%", "97.5%"))

Theta2_n4<-matrix(c(colMeans(t(X4))[2], Varianza_renglones(X4)[2],
I2_n4), 1, 4)
dimnames(Theta2_n4)=list(c(n4),
                          c("Media","Varianza", "25%", "97.5%"))

Theta3_n4<-matrix(c(colMeans(t(X4))[3], Varianza_renglones(X4)[3],
I3_n4), 1, 4)
dimnames(Theta3_n4)=list(c(n4),
                          c("Media","Varianza", "25%", "97.5%"))

L1<-rbind(Theta1_n1, Theta1_n2,Theta1_n3, Theta1_n4)
L2<-rbind(Theta2_n1, Theta2_n2,Theta2_n3, Theta2_n4)
L3<-rbind(Theta3_n1, Theta3_n2,Theta3_n3, Theta3_n4)

Final<-list(L1,L2,L3)
names(Final)<-c("Theta1", "Theta2", "Theta3")

return(Final)
}
#Vasicek_media_varianza_intervalos(1000,1,2,0.5,.5,1000,500,200,100)

```

Argumentos

- **N** Número de pasos a simular.
- **theta1** parámetro del proceso .
- **theta2** parámetro del proceso.
- **theta3** parámetro del proceso .
- **delta** Tiempo entre cada paso de la simulación.
- **n1** número de trayectorias a simular.

- **n2** número de trayectorias a simular $\leq n1$.
- **n3** número de trayectorias a simular $\leq n1$.
- **n4** número de trayectorias a simular $\leq n1$.

A.2. Estimación máxima verosímil por el método de Euler-Maruyama

A.2.1. Estimación para el proceso de Ornstein-Uhlenbeck

Descripción

Estimación_EM_Ornstein() hace uso de la paqueteíara “sde”, simula una trayectoria del proceso de Ornstein-Uhlenbeck y estima los parámetros mediante el algoritmo descrito en la subsección 7.1.1.

Código

```
library(sde)

Estimacion_EM_Ornstein<-function(N, theta1, theta2, theta3, deltat,
                                X0){
  X <- sde.sim(model="OU", theta=c(theta1,theta2,theta3), N=N,
              delta=deltat, X0=X0)
  n<-length(X)
  theta.gorro<-(sum(X[1:(n-1)]^2)+1/n*(sum(X[1:(n-1)]*(X[n]-X[1])))-
               sum(X[1:(n-1)]*X[2:n]))/
  (deltat*(sum(X[1:(n-1)]^2)-1/n*(sum(X[1:(n-1)]))^2))
  m.gorro<-(sum(X[1:(n-1)])-sum(X[2:n])+theta.gorro*deltat
            *sum(X[2:n]))/(n*theta.gorro*deltat)

  m.gorro.gorro<-m.gorro*theta.gorro
  sigma.gorro<-sqrt((sum((X[1:(n-1)]-X[2:n]-
                        (theta.gorro*(m.gorro.gorro-X[2:n])
                        *deltat))^2))/(n*deltat))
  estimadores<-data.frame(m.gorro.gorro,theta.gorro,sigma.gorro)

  colnames(estimadores)<-c("theta1", "theta2", "theta3")

  return(estimadores)
}
```

Argumentos

- **N** Número de pasos a simular.
- **theta1** parámetro del proceso.
- **theta2** parámetro del proceso.
- **theta3** parámetro del proceso.

- **deltat** Tiempo entre cada paso de la simulación.
- **X0** valor inicial del proceso.

A.2.2. Estimación para el proceso de Cox-Ingersoll-Ross

Descripción

Estimación_EM_CIR() hace uso de la paquetería “sde”, simula una trayectoria del proceso de Cox-Ingersoll-Rossy estima los parámetros mediante el algoritmo descrito en la subsección 7.2.2.

Código

```
library(sde)

Estimacion_EM_CIR<-function(N, theta1, theta2, theta3, deltat, X0){
  X <- sde.sim(model="CIR", theta=c(theta1,theta2,theta3), N=N,
              delta=deltat, X0=X0)
  n<-length(X)

  kappa.gorro<-(n^2-n*sum((X[2:n])/X[1:(n-1)]))+
    sum((X[n]-X[1])/X[1:(n-1)]))/
    (deltat*(n^2-sum(X[2:n])*sum(1/X[2:n])))

  alfa.gorro<-(sum(X[2:n]/X[1:(n-1)])-n+n*deltat*kappa.gorro)/
    (deltat*kappa.gorro*sum(1/X[1:(n-1)]))

  sigma.gorro<-sqrt((sum((1/X[1:(n-1)])*
    (X[2:n]-X[1:(n-1)]-
    kappa.gorro*(alfa.gorro-X[1:(n-1)])*
    deltat)^2))/(deltat*n))

  estimadores<-data.frame(kappa.gorro, alfa.gorro*kappa.gorro,
    sigma.gorro)

  colnames(estimadores)<-c("alfa", "kappa", "sigma")

  return(estimadores)
}
```

Argumentos

- **N** Número de pasos a simular.
- **theta1** parámetro del proceso.
- **theta2** parámetro del proceso
- **theta3** parámetro del proceso.

- **deltat** Tiempo entre cada paso de la simulación.
- **X0** valor inicial del proceso.

Bibliografía

- [1] Bernt Oksendal *Stochastic Differential Equations An Introduction with Applications*. Springer-Verlag Heidelberg New York 1997.
- [2] I.I. Gihman ,A.V. Skorohod *Stochastic Differential Equations*. Springer-Verlag, Nueva York Heidelberg Berlin, 1972.
- [3] Stefano M.Iacus. *Simulation and Inference for Stochastic Differential Equations*. Springer, 2008.
- [4] Peter E.Kloeden, Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*. Springer 1991.
- [5] Yacine Aït-Sahalia *Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach*. *Econometrica*, Vol. 70, No. 1 (January, 2002), 223-262.
- [6] Ioannis Karatzas Steven E. Shreve *Brownian Motion and Stochastic Calculus*. Springer 1991.
- [7] Rubens Penha Cysne *On the Statistical Estimation of Diffusion Processes: A Partial Survey*. *Brazilian Review of Econometrics* 2004.
- [8] Søren Asmussen *Stochastic simulation Algorithms and Analysis*. Springer 2000.
- [9] Sorensen H. *ametricinferencefor diffusion processes observed at discrete points in time: A survey*. University of Copenhagen 2002.
- [10] Søren Asmussen *Stochastic simulation Algorithms and Analysis*. Springer 2000.
- [11] K. Itô *Differential equations determining Markov processes*. Zenkoku Shijō Suga-ku Danwakai 1942.